

## CHAPTER 4

# Breaking Out of the Physicalist Triad

In the previous three chapters I outlined three philosophical positions that I believe are mutually reinforcing: (1) physicalism, (2) causal theories of intentional action and (3) relational approaches to causation. I have called this triad of views *the physicalist triad* because the consequence of endorsing each element of the triad is that physicalism about mentality becomes the only acceptable metaphysics of mind: it appears to be the only view that has a chance of saving the phenomenon of mental causation. In many arguments for physicalism, mental causation is understood in relational terms; that is, mental causation is presented as a cause–effect relation between mental and physical items. 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—so the only thing that can discriminate between different categories of causation is the nature of the relata involved. This assumption about causation, an assumption I have called ‘relationalism’, is ubiquitous in philosophy of causation but it is also a thesis that will be appealing to physicalists because of its associations with naturalism. Mental causation is also made to seem indispensable because of causal theories of intentional action. Causal theories of intentional action, however, owe their dominance to relational assumptions about causation. This is because, if causation is always, everywhere a relation, then explaining what intentional action is is a matter of distinguishing between different types of event causation (those that do and those that do not constitute intentional action). In summary, then, even though each element of the triad is logically independent, in practice they reinforce each other. Physicalists endorse relational approaches to causation because they are naturalistic; against the backdrop of the relational approach to causation, causal theories of intentional action are made to seem intuitively more appealing than their rivals; and endorsing causal theories of intentional action strengthens the case for physicalism by making relationally understood mental causation seem indispensable.

---

**How to cite this book chapter:**

White, Andrea. 2024. *Understanding Mental Causation*. Pp. 73–96. York: White Rose University Press. DOI: <https://doi.org/10.22599/White.e>. License: CC BY-NC 4.0

In this chapter, I want to explain why I think we should try to break out of the physicalist triad. Physicalism, causal theories of intentional action and relational approaches to causation are individually plausible, with a lot of explanatory power. Physicalism purports to offer a naturalistic account of the mind; causal theories of intentional action promise to explain how it is that reasons can explain actions as well as offering an account of what makes the difference between what an agent does and what happens to him; relational approaches to causation, such as the regularity theories, counterfactual theories and manipulability theories are powerful theories about what causation is. Given how well supported each element of the physicalist triad is, if I have shown that they are also mutually reinforcing, perhaps this is just another reason to favour them. So why do I think we should try to break out of the physicalist triad?

The weakest point of the triad, or so it seems to me, is the account of agency it provides. The physicalist triad entails a physicalist/event-causalist description of agency and, as I explain in this chapter, this description of agency faces a number of problems. First, there is the problem presented by apparent counterexamples that involve deviant causal chains from mental cause to bodily movement. Second, there is the difficulty posed by the fact that sometimes agency is manifested through refrainment, i.e. by not doing anything. Third, there is the problem of giving an account of actions that are less than fully intentional. These problems will be familiar to anyone keeping track of debates within philosophy of action. However, I will argue that these problems are not three distinct issues. Instead they are all symptoms of a more fundamental issue with a physicalist/event-causalist description of agency, which is the assumption that the distinction between ‘agential’ and ‘non-agential’ can be understood in terms of a distinction between different kinds of event-causal sequence.

#### 4.1 The disappearing agent

Very generally, agency refers to the power to act. Part of the task of philosophy of action is to explain what it is to act. The physicalist/event-causalist answer to this question construes what it is to act in terms of intentionality: what it is to act is to do something intentionally, which entails that all actions are intentional under some description. Davidson argued for this position, claiming that ‘a man is the agent of an act if what he does can be described under an aspect that makes it intentional’ (1971/2001a: 46).

What it is for an action to be intentional is then explained in terms of causation by a mental state of the agent, or a mental event involving the agent (this is the causal theory of action). The difference between a bodily movement that is intentional and one that it is not ‘lies in the causal aetiology of what happens when a body moves’ (Smith 2012: 387). 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 Alfred Mele (1992a; 2003), who, in his own words, defends ‘a causal perspective on intentional action’ which consists of a pair of theses: ‘(1) all intentional actions are *caused* (but not necessarily deterministically so); (2) in the case of any intentional action, a causal explanation framed part in terms of *mental* items (events or states), including motivation-encompassing attitudes, is in principle available’ (Mele 2003: 5). Mele says the second thesis can be developed by adding that ‘the relevant mental items are realised in physical states and events that are important causes of intentional actions, and—owing to the particular relations of the mental items to the realising physical items, to appropriate counterfactual connections between the mental items and the actions, and to the truth of relevant psychological and psychophysical generalisations—the mental items properly enter into causal explanations of the actions’ (2003: 5). Mele has also defended a ‘causal approach to analysing and explaining actions’, which he describes as the view that ‘our actions are, essentially, events (and sometimes states, perhaps) that are suitably caused by appropriate mental items, or neural realisations of those items’ (2000: 279). So, although Mele is focused primarily on defending the viability of a causal account of what intentional action is, he is at least open to the possibility that such an account could be given a physicalist development.

Berent Enç also defends the causal theory of action, which he describes as ‘the proposition that an act consists of a behavioural output that is caused by the reasons the agent has for producing that behaviour—reasons that consist of the beliefs and desires of the agent’ (2003: 2). As stated in Chapter 2, Enç believes that deliberation is a ‘computational process’ that results in an intention that in turn causes an item of behaviour. ‘On this thesis,’ states Enç, ‘actions are defined as changes in the world that are caused by mental events.’ Enç also states that he ‘[helps himself] to the assumption that mental attributes like beliefs, desires, hopes, value judgements and so forth are manifestations of the physical world, and that what is generally referred to as naturalism is the correct view about such mental events and states’ (2003: 2). The physicalist/event-causal description of human agency is also defended, in whole or in part, by Brand (1984), Bishop (1989), Bratman (1987), Dretske (1988), and recently by Shepherd (2021).

Several arguments presented in philosophy of action appear to show that seeking to understand agency in terms of a distinction between different types of event causation cannot be done without misconstruing the agency concept. For example, Jennifer Hornsby (2004) argues that the physicalist/event-causalist description of agency ‘leaves agents out,’ which is problematic because ‘human beings are ineliminable from any account of their agency’ (2004: 2).

This objection has come to be known as ‘the disappearing agent problem.’ According to this objection, an essential part of our concept of agency is that,

in acting, the agent herself brings about changes. However, as David Velleman (1992: 461) puts it, causal theories of intentional action entail that the agent is ‘merely the arena’ within which mental states or events cause bodily movements. The agent herself does not bring about what she intends. In this way, the agent ‘disappears’. This cannot be right because a world where agents do not bring about the results of their actions is a world where there are no actions. This objection is, I believe, devastating to the physicalist/event-causalist description of human agency. However, it is often misunderstood.

One way it is misunderstood is to see it as begging the question against the event-causal theory of action. All versions of the event-causal theory of action hold that acting intentionally consists in the right kind of event being caused to happen, in the right way, by the right kind of mental antecedents. The core proposal of the event-causal theory is that acting intentionally is nothing over and above some special kind of event causation. The disappearing agent problem can seem like a straightforward denial of the event-causal theory’s core proposal. The critic of the event-causal theory complains that the agent is missing from an account of her agency, while the event-causal theory’s core thesis is that mental states causing bodily movements *is* the agent carrying out her agency. In her summary of the disappearing agent objection, Sarah Paul characterises the disappearing agent objection as committing a category mistake:

The complaint is sometimes put in terms of the subject being a ‘mere arena’ in which psychological states are contained, such that she is not involved in the interactions between mind and body. But the Causal Theorist is in no way committed to this way of thinking about the relationship between the subject and her own mind. Indeed, this seems to be a prime example of a category mistake: ‘I see that there are mental states, and a body that moves around in virtue of this mental activity, but where is the person that does the moving?’ (2020: 56)

If the disappearing agent problem is understood this way, then event-causal theorists can respond by insisting that the agent does not disappear on their account because the agent’s bringing about what she intends is identical with mental states of the agent causing bodily movements.

Another way the disappearing agent problem is misunderstood is to see it as revealing that the standard version of the causal theory of action—i.e. the version which says that an intentional action is a bodily movement which is caused by an intention to act, which is in its turn caused by desire for something and a belief about how to satisfy that desire—is insufficient to capture intentional agency. This is how Velleman (1992) understands the problem. Velleman argues that in the standard version of the causal theory of action there is nothing—no mental state, or causal sequence—that amounts to the agent taking an active part in her action. However, for Velleman, this does not show that no version of the causal theory of action can succeed. Velleman thinks the

disappearing agent problem shows that the causal theory of action needs to be modified but not rejected.

Velleman argues that the standard version of the causal theory of action actually succeeds as an account of what it is to act ‘half-heartedly, or unwittingly, or in some equally defective way’ (1992: 462). That is, the standard version of causal theory of action does capture *a kind* of action, but it ‘describes an action from which the distinctively human feature is missing ... not a human action par excellence’ (1992: 162). Velleman’s opinion is that sub-par action, which he describes as ‘half-hearted’, ‘unwitting’, ‘defective’, consists of mental states like desire, belief and intention taking our bodies from inactivity to activity. In cases of sub-par action, the flux of events—which includes mental events—operates through us but does not involve us—we play no active part. In human action par excellence, by contrast, we are involved and do play an active part.

In a full-blooded action, an intention is formed by the agent himself, not by his reasons for acting. Reasons affect his intention by influencing him to form it, but they thus affect his intention by affecting him first. And the agent then moves his limbs in execution of his intention; his intention doesn’t move his limbs by itself. The agent thus has at least two roles to play: he forms an intention under the influence of reasons for acting, and he produces behaviour pursuant to that intention. (1992: 462)

According to Velleman, the active part we play can be reduced to the causal role of some mental state of ours. Specifically, ‘a motive that drives the agent’s critical reflection on, and endorsement or rejection of, the potential determinants of his behaviour, always doing so from a position of independence from the objects of review’ plays the functional role of the agent in action par excellence (1992: 477). As long as this higher-order motive is included in the event-causal story leading up to a bodily movement, the causal sequence described amounts to action par excellence. If the disappearing agent objection is understood as merely showing that the causal theory of action needs to be modified, then it does not disprove the physicalist/event-causal description of agency, nor does it give us a reason to break out of the physicalist triad.

The third way the disappearing agent problem is misunderstood is to see it as a problem for event-causal accounts of *a special kind* of action, as opposed to action in general. A key example of this kind of misunderstanding can be seen in Derk Pereboom’s (2014) argument for understanding *free will* in terms of agent causation.

Pereboom argues that event-causal libertarian theories of free will are inadequate. Libertarians about free will believe that an action cannot be free if it is deterministically caused to happen by a prior event (incompatibilism). Pereboom argues that simply injecting indeterminism into the causal chain leading up to an action cannot secure freedom. This is because ‘if only events are causes and the context is indeterministic, the agent disappears when it

needs to be settled whether the [action] will occur' (2014: 55). The point here is that, in a determinist event-causal sequence, prior events 'settle' whether an action occurs, therefore, for the incompatibilist, the action cannot be free, but in an indeterministic system *nothing* settles whether the action occurs, and for *that* reason the action cannot be free. Free action requires that *the agent* settles whether the action occurs or not. In an event-causal system, even one which involves indeterminacy, the agent is not settling anything—they have disappeared—and so free actions do not exist. The solution is to hold that an action is free just in case it is caused to happen by the agent. Now the agent, rather than any prior event, is the causal determiner of the action. The thought is that agent causation best captures the sense in which free agents need to, themselves, be the settlers of their actions.

Pereboom's argument fails as an argument against event-causal libertarian theories of free will. As Randolph Clarke (2017) argues, the event-causal libertarian can grant Pereboom's condition that an action is only free if the agent (and no prior event) settles whether the action occurs or not but insist that this condition is met on her account of free action. The event-causal libertarian can argue that whether the action occurs or not is settled by the agent when the action occurs. Prior to the agent's action it is an open question whether the action will occur or not. The matter is not settled prior to the agent's action because the events that cause the action do not deterministically cause the action. However, when the action occurs, the question of whether the action will occur or not is closed, and thereby settled. The occurrence of the action itself settles whether the action occurs. Therefore it is not the case that *nothing* settles whether the action occurs. The event-causal libertarian can argue that the occurrence of the agent's action *is* the agent's settling of whether the action occurs or not.

One might defend Pereboom's claim that, if only events are causes and the context is indeterministic, then the agent does not settle anything by insisting that the event-causal libertarian fails to specify conditions that are sufficient for the agent to do anything at all. If the agent does not act, then no event could constitute the agent's settling of something. However, this would change the target of Pereboom's argument. Pereboom explicitly accepts that it is still possible for agents to act even if only events are causes; he only argues that none of these actions can be free. To argue that the event-causal libertarian fails to specify conditions that are sufficient for the agent to do anything at all is a different argument. Pereboom's argument fails, I think, because the disappearing agent objection is really an issue about the possibility of action itself, it is not specifically to do with freedom.

I have presented three ways in which the disappearing agent should *not* be understood, so how *should* we understand this problem? The disappearing agent problem is not best expressed as a direct challenge to causal theories of action. So expressed, it can seem like it is begging the question. Furthermore, the disappearing agent problem is not about a special kind of action, e.g. action

par excellence or free action. The essence of the disappearing agent problem is that our general concept of agency is *fundamentally at odds* with a view of the world that assumes that causal reality is nothing but a chain of causally related events, a worldview where ‘all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another’ (Lewis 1986: ix). Perhaps the best expression of the problem comes from Abraham Melden:

It is futile to attempt to explain conduct through the causal efficacy of desire—all that can explain is further happenings, not actions performed by agents. The agent confronting the causal nexus in which such happenings occur is a helpless victim of all that occurs in and to him. There is no place in this picture of the proceedings either for rational appetite or desires, or even for the conduct that was to have been explained by reference to them. (1961: 128–129)

Melden describes the aim of theories like the causal theory of action as ‘futile’. He thinks that no event-causal theory of action could succeed; such a theory will always fail to adequately capture our thinking about agency. Melden claims that within ‘the causal nexus’ the agent becomes ‘a helpless victim of all that occurs in and to him’.

This claim needs a bit of explaining. The issue 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; the agent seems passive, like a victim. This metaphor of the agent becoming a ‘victim’ is why, I think, the disappearing agent problem can seem like it is about free action, or action par excellence, but it is important not to get carried away by the metaphor. The essential point is 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 do not marry: agency is about agents doing things—a causally related chain of events contains only what occurs or happens.

Thomas Nagel (1986) also expresses the disappearing agent problem well. For Nagel, part of the problem with the physicalist/event-causalist picture of agency is that there are important truths about agency that are lost when we view the causal world from a physicalist/event-causal perspective. On the physicalist/event-causal picture, the causal world is a ‘spatiotemporal mosaic’ of instantiations of categorical, objective properties (Lewis 1994: 474) or ‘the fusion of all events throughout space-time’ (Schaffer 2007: 83). According to Nagel, ‘something peculiar’ happens when we attempt to describe action from this ‘objective or external standpoint’.

Actions seem no longer assignable to individual agents as sources, but become instead components of the flux of events in the world of which the agent is a part ... There seems no room for agency in a world of neural impulses, chemical reactions, and bone and muscle movements. Even if we add sensations, perceptions, and feelings we don't get action, or doing—there is only what happens. (1986: 110–111)

Should the disappearing agent problem be taken seriously? Is our general concept of agency really *fundamentally at odds* with a view of the world that assumes that causal reality is nothing but a chain of causally related events? Is it really 'futile' to try to explain what it is to act in terms of causation of bodily movements by mental events? I think we should answer these questions affirmatively and, for me, this is the main motivation for breaking out of the physicalist triad. However, this is a bold claim and I will need to defend it.

The disappearing agent problem should be taken seriously because the physicalist/event-causal picture of agency fails in three important ways: it fails to solve the problem of deviant causal chain cases; it fails to account for refrainment; and it fails to account for the unity between intentional agency and non-intentional agency. The best explanation for these failures is because the physicalist/event-causal picture of agency leaves no room for the agent.

#### 4.1.1 Deviant causal chains

The physicalist/event-causal picture of agency construes what it is to act in terms of intentionality: what it is to act is to do something intentionally. The causal theory of action says that intentional actions are bodily movements caused, in the right way, by certain mental states of the agent or mental events involving the agent. In Chapter 2, I mentioned that the most significant source of disagreement about how the causal theory of action should be formulated concerns what constitutes *the right way* for a mental state or event to cause a bodily movement for there to be intentional action. Not just any causal chain from mental event to physical event is sufficient for there to be an intentional action. These mental states have to operate in the causal chain in the right way. A necessary condition for acting intentionally is that the agent is in control of what is going on with them. For there to be intentional action, the causal chain from mental item to bodily movement must be such that it constitutes the agent's control over their action. The causal chain cannot deviate from the kind of causal chain that occurs in a normal, uncontroversial case of intentional action. Davidson gives an example of a deviant kind of causal chain:

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, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally. (Davidson 1973/2001a: 79)

In this example, the climber has an end he wants to achieve, namely to rid himself of the weight and danger of holding the other man, and the climber reasons that loosening his hold is the best means to achieve this end. This belief–desire pair causes a bodily movement of a type that is rationalised by the belief–desire pair, just as causal theorists allege it would in an ordinary case of intentional action. But in this case the causal route from belief–desire pair to bodily movement involves an intermediary state of nervousness that ‘robs the climber of control’, as John Bishop (1989: 134) puts it. In this example, the climber did not let go intentionally. The challenge for the causalist that deviant causal chains present is to ‘specify the sorts of causal paths that can count as the “right” way in which beliefs and desires must yield behaviour for genuine intentional action to occur’ (Bishop 1989: 135).

There is great disagreement on what kind of causal chain from mental state to bodily movement is required for the agent to retain control over their action. Davidson himself doubted that a reductive analysis of intentional action could be developed from the idea that states of desiring and states of believing are causes of the actions they explain because of deviant causal chain cases. However, many have argued that a reductive causal analysis of intentional action is still possible, Davidson’s nervous climber example notwithstanding. Davidson’s example shows that the original causal theory failed to specify jointly sufficient necessary conditions for intentional action, but this does not mean that a more sophisticated version of the causal theory will also fail. Many more sophisticated versions of the causal analysis of intentional action have been offered since 1973.

One promising strategy is the ‘sensitivity approach’ (e.g. Bishop 1989; Mele 1992a; Mele 2003; Peacocke 1979). This approach suggests that a necessary condition for intentional action is that the bodily movement caused by the relevant mental state is ‘responsive’ or ‘sensitive’ to the content of the mental state. One way of spelling out this sensitivity requirement is in terms of counterfactuals: a bodily movement is sensitive to the mental state that caused it if and only if a slightly different bodily movement—one that conformed to the different mental state—would have occurred had the agent’s mental state had a slightly different content.<sup>15</sup> Smith (2010) gives a clear example: suppose a pianist wants

---

<sup>15</sup> The counterfactual version of the sensitivity approach isn’t the only version available. Peacocke (1979: 69) offers an alternative version. Peacocke argues that there is an intentional action if and only if the bodily movement is caused by an intention and that the intention differentially explains the occurrence of the bodily movement. A state or event differentially explains

to appear nervous to his audience and believes he can achieve this end by playing a C# instead of a C during his piece. The pianist's pressing C# is sensitive to this belief–desire pair if and only if the pianist would have pressed B had he thought that pressing B would achieve his goal. Cases of deviant causation are thought not to satisfy this sensitivity requirement.

However, this proposal faces a counterexample. Consider again my friend Amy, who has a device that can manipulate my brain and nervous system in the manner of the character Black from Harry Frankfurt's (1969) thought experiment. Amy can use this device to control my bodily movements as an engineer might control a remote-operated machine. When Amy uses her device, what happens with my body is not up to me; I am not in control of my bodily movements and therefore am not demonstrating agency. When Amy uses her device, she has taken control over what goes on with me. Now suppose that Amy uses her device to move my body to carry out my own intentions. For example, suppose I form the intention to make tea, and in response to this Amy uses her device to make me make tea. Suppose further that had I formed a different intention Amy would have used her device to make sure my body moved in conformity with my alternative intention.<sup>16</sup> In this strange example, the bodily movement that results from my intention to make tea is sensitive to the content of that intention. However, when Amy uses her device to manipulate my brain and nervous system, I am not performing an intentional action: I am not in control over what is going on with my body; Amy is. Bishop calls cases like this, where the causal path from intention to bodily movement passes through a benevolent second agent, 'heteromesial' causal chain cases.

A more recent suggested solution to the causal deviance problem, proposed by McDonnell (2015), also cannot deal with this counterexample. McDonnell suggests that there is an intentional action if and only if the mental cause of the bodily movement is 'proportional'; in Stephen Yablo's (1992) sense, to the bodily movement. My intention to make tea is a proportional cause of my subsequent tea-making if and only if the following counterfactual conditionals are true:

1. Had my intention to make tea been absent, then I would not have made tea.
2. Had my intention to make tea been absent, then had I intended to make tea I would have made tea.

---

another when there is a law backing the explanation, according to which changes in the intensity or value of the explanandum are correlated (one-to-one) with changes in the intensity or value of the explanans. For the sake of brevity, I won't discuss Peacocke's version of the sensitivity approach here. See Sehon (1997) for a convincing argument that Peacocke's proposed criterion for intentional action is neither necessary nor sufficient.

<sup>16</sup> This counterexample is adapted from an example given by Peacocke (1979: 87).

These are both true even in the heteromesial case.

One obvious response to such cases is to stipulate that the causal chain cannot be heteromesial if intentional action is to occur. However, as Bishop points out, this cannot be right, as not every heteromesial causal chain is such that it blocks intentional action. Bishop (1989: 125) describes a case where machinery like Amy's is used to make sure that damaged neural pathways carry on functioning as normal (e.g. suppose some synapse isn't functioning properly; Amy's machinery might work by stimulating the second neurone when the first is in the right electrochemical state, just as the first neuron would if it were working properly). Even if Amy had to hold a switch down to keep the machinery working, so that the causal chain from intention to bodily movement must go via an action of Amy's, this would not necessarily mean that no intentional action is possible in this case. Suppose I'm the one with the damaged neural pathways, and Amy has to keep the machine switched on when I decide to make tea. In this case Amy is helping me carry out my intention to make tea by helping my nervous system remain in working order—she's an essential component of the causal chain that lets me carry out my intention, but it is less clear that I lack agential control in this case. It is not the involvement of a second agent *per se* that is adversative to intentional action but the manner of their involvement.

The problem posed by deviant causal chain cases may be solvable. It might be possible to give a counterexample-free specification of what constitutes the right path from mental cause to bodily movement for the bodily movement to count as an intentional action. On the other hand, the project of specifying what it is for a causal chain from intentions to bodily movement to be non-deviant may suffer a similar plight to that faced by the project of specifying necessary and sufficient conditions for knowledge, namely that every new proposal faces new counterexamples and the project seems nowhere near an end.<sup>17</sup> The pessimistic conclusion is that deviant causal chain cases should make us doubt that causation by a mental event constitutes what it is to act. In *some* cases, causation by a mental event seems to put the mental condition of a person in control of a bodily movement *at the expense of* the person themselves. If we cannot distinguish such cases from genuine cases of agency in event-causal terms, then the idea that personal control over one's body consists in causation of a bodily movement by a mental state is doubtful.

#### 4.1.2 Refrainment

Sometimes human beings demonstrate their agency by not acting. For example, imagine I let a plant die by not watering it. This is an example of refraining from acting and thereby allowing something to happen. Other examples of

---

<sup>17</sup> See Zagzebski (1994) for an argument that Gettier-style counterexamples are inescapable for almost every analysis of knowledge.

refrainment include offending someone by not greeting them (Alvarez 2013: 104) or allowing a telephone to continue ringing by not answering it (Hornsby 2004: 5). Another interesting case comes from John Hyman (2015: 10–11). Hyman uses an example of a child being picked up by a parent to show that sometimes passivity is voluntary. With respect to being picked up, the child is passive, but being picked up is voluntary for the child. We know being picked up is voluntary for the child because the child could resist (e.g. by pushing away the parent or crying) but does not. We can further suppose that the child wants to be picked up and does not resist because she wants to be picked up. This qualifies the case as an instance of intentional passivity. In this case, the child is demonstrating an agential power, even though the child is, so to speak, not doing anything but, rather, letting something happen to her. The child is demonstrating agency by not resisting. In this case, there is an action, but it is the action of the parent not the child. In these examples, what occurs is at least partly up to the agent—the agents have that kind of control over what happens. This suggests that these examples are examples of agency. However, there are no actions in these examples, so their status as agential cannot be explained as the causation of an action by a mental state or event.

Bruce Vermazen (1985) describes a subclass of actions called ‘negative acts’. One could challenge the claim that there are no actions in the examples above by arguing that in the examples the agents perform negative actions. However, I do not think the above examples are correctly described as negative acts. They are instead what Randolph Clarke (2014) calls ‘omissions’, which he argues are the absences of action. Maria Alvarez (2013) describes an example of refrainment that I think illustrates what a ‘negative act’ is, even though Alvarez would not herself describe the example as such. In Alvarez’s example, an agent stands motionless in front of a laser-beam mechanism that controls a door and thereby prevents the door from closing by not moving. In this example, there is no positive performance by the agent. However, there does seem to be something that the agent does. Standing motionless seems to be an action, albeit one that is described in negative terms. In contrast, in the plant example, watering my plant is an action available to me that I simply do not do, thereby I allow other events (transpiration perhaps) to cause the plant to die. In the example where I offend someone by not greeting them, it is the absence of an act of greeting that matters. Similarly, in the telephone example, not answering the telephone is not an action negatively described but the absence of an action.

What makes refrainment a demonstration of agency? The causal theory of action is not equipped to answer this question. If being capable of agency is just to possess mental states that cause actions to happen, as the causal theory of action proposes, then it should be impossible for people to demonstrate agency when they do not perform an action. Because refraining is not acting, what makes refrainment a demonstration of agency cannot be expressed in terms of the mental causation of an action. However, examples of refrainment are not counterexamples to the causal theory of action. The causal theory of action is

only an account of *action*; it does not purport to explain what refrainment is. What these examples indicate is that the causal theory of action cannot tell the whole story about agency in terms of causation of an action by a mental event. To give a full explanation of what agency is, we need to explain why agency can be manifested not only by performing an action but also by refraining from acting, a fact that initially seems very puzzling given that agency is the power to act.

The important question is whether it is possible to explain refrainment in a way that abides by the physicalist and relationalist assumptions of the physicalist/event-causal picture of agency. Relationalism says that causation is always and everywhere a relation between distinct entities ('cause' and 'effect') that are normally supposed to be events. There may be some theories of what events are that allow something's not-happening to count as an event,<sup>18</sup> but on any theory that takes seriously the idea that events are happenings, this proposal that omissions are events is implausible: something's not-happening is not a thing that happens.<sup>19</sup> Clarke (2014) argues that omissions are non-entities; that is, they are not things that exist at all. Relationalism thus seems to rule out that omissions could be causes or effects. This appears to rule out any event-causal explanation of refrainment: the agency of refrainment cannot consist in omissions being caused to happen by mental states or events because omissions cannot happen.

Clarke (2014), however, offers an account of refrainment that appears to be compatible with relationalism about causation. Clarke argues that his account of refrainments is compatible with the view that omissions cannot be causes or effects as they are non-entities. Clarke argues that, 'in a case of intentional omission or refraining, relevant mental states (or events) must cause some of the agent's subsequent thought or conduct, even if they needn't cause the absence of some action' (2014: 75). As he puts it elsewhere, 'in cases of intentionally omitting or refraining, some intention with relevant content must play a causal role with respect to some of what subsequently does happen—with respect to one's subsequent thought and conduct' (2014: 78). For example, suppose we accepted that the child's desire to be picked up could not be the cause of her not resisting because not resisting is an absence and therefore cannot be an effect. Clarke's account of refrainment allows us to explain the intentionality of this omission as consisting in the child's desire causing some of the child's subsequent behaviour. Suppose wanting to be picked up caused the child to put her arms around the parents shoulders (and thereby make her being picked up easier)—that would be what makes the child's not-resisting intentional,

---

<sup>18</sup> For example, on certain theories of events as property exemplifications it might be possible for there to be negative events. Philosophers who have argued for the reality of negative events include: De Swart (1996), Higginbotham (2000) and Vermazen (1985).

<sup>19</sup> See Mele (2005) for further reasons to reject negative events.

according to Clarke's proposal. If Clarke's proposal succeeds, then the physicalist/event-causal account of agency only needs to be slightly amended to include refrainment in its account of agency. The amended account would be: what it is to demonstrate agency is to do *or not do* something intentionally and what it is for an action *or omission* to be intentional is explained in terms of causation by a mental state of the agent, or a mental event involving the agent.

Although Clarke's account of intentional omissions is similar to the causal theory of intentional action insofar as mental causation is an essential part of what makes an omission intentional, some of what Clarke says about intentional omissions is anti-relationalist in spirit. The intentionality of omissions, and hence the agency of omissions, does not consist in their being caused to happen by any event. Instead, what makes an omission intentional is that it sits within a larger sequence of thoughts and actions that demonstrates a teleological structure. To find what makes omissions intentional we must look at the wider context of the agent's behaviour. The intentionality of the omission is not revealed if we consider the omission in isolation. Instead we have to see the omission as part of a larger pattern of activity that is directed towards an end that is incompatible with performing the omitted act. Even if mental causation is essential for understanding the intentionality of refrainment, these cases lend support for the idea that the physicalist/event-causal picture of agency cannot be the whole story about agency.

#### 4.1.3 *Over-mentalisation of agency*

The third important failure of the physicalist/event-causal picture of agency concerns its treatment of agency that is less than fully intentional. On the physicalist/event-causalist view, to act is to do something intentionally. Agency is thus explained in terms of intentionality. However, not all examples of agency are also examples of intentional action.

In what follows, I will give three examples of agency that lack the typical characteristics of intentional action. Most proponents of the physicalist/event-causal picture of agency assume that the typical characteristics of intentional actions are as follows:

- (a) they are done for reasons, which is to say that there is a true description of the action which makes the action seem to the agent to be a sensible or rational or good thing to do;
- (b) they are done in order to achieve a goal, which is to say that there is some further action that the agent is trying to complete and her intentional action is, she believes, a means by which she can complete that further action;
- (c) they are subject to rationalising explanations, which is to say that they can be causally explained by facts about what the agent wants to do and

facts about what the agent believes about how to do it (note that, on the physicalist/event-causalist view, rationalising explanations are causal explanations).

Of course, I acknowledge that there are theories of intentionality where actions can be intentional even though they lack some, perhaps even all of these features. Therefore, there may be conceptions of intentionality under which the examples I give here *do* count as intentional. (Indeed, some of these examples may count as intentional on the conception of intentional action which I propose in Chapter 9.) However, as the purpose of this chapter is to challenge the physicalist/event-causal picture of agency, what matters dialectically is whether the examples I describe in this section conform to the physicalist/event-causalist characterisation of intentional action. If they do not, then the physicalist/event-causalist idea that to act is to do something intentionally is under pressure.

The first examples of agency that lack the typical characteristics of intentional action are actions that Brian O'Shaughnessy calls 'sub-intentional'. Sub-intentional actions include actions like 'tapping my feet to the music' and 'idly moving my tongue in my mouth' (1980: 61), actions we'd often describe as 'absent-minded'. Other examples may include shifting one's position, automatically scratching an itch or fiddling with one's hair.

Sub-intentional actions are not actions that seem, to the agent at the time of performing them, like sensible, or rational or good things to do, nor are they actions performed in pursuit of a goal. Sub-intentional actions also cannot be rationalised by facts about what the agent wants to do and what the agent believes about how to do it. Actions like tapping one's foot to music or shifting one's position, or fiddling with one's hair do not seem to be preceded by or accompanied by (and hence not causally explained by) an intentional state such as believing that performing the action is a good idea, or wanting to achieve something by means of the action. When I absent-mindedly tap my feet to the music, it is not true that I do this because there is something I want to do and believe that tapping my feet is a means by which I can do it. Furthermore, at the time of performing a sub-intentional action, the agent is often not aware that she is performing the action at all. The actions O'Shaughnessy delineates are actions about which we'd often say "Oh, I didn't realise I was doing that." For these reasons, sub-intentional actions, despite being under my control, seem to lack characteristics (a)–(c).

Sub-intentional actions also do not seem to be the causal consequence of an episode of thought. O'Shaughnessy thinks that sub-intentional actions are subject to psychological explanations. For example, he suggests that sub-intentional actions might be explained in terms of feelings of restlessness (1980: 61). When I shift my position, it is usually because I feel uncomfortable. I might fiddle with my hair because the sensation is comforting to me. I concede that sub-intentional actions can be explained in terms of feelings or sensations.

However, these psychological explanations do not seem to point to or mention a specific mental event that preceded the action and which could be considered the cause of the action. The explanations seem to cite concurrent experiences, as opposed to episodes in the agent's mental history, which caused her to fidget or fiddle. For this reason, sub-intentional actions seem to be exercises of agential power which do not have mental causes.

Another important class of human actions which lack the typical characteristics of intentional actions are *spontaneous expressions of emotion*. Examples include embracing a loved one, crying upon hearing bad news, laughing at a joke, wincing when you make a mistake, or shouting at your computer after it crashes at an inconvenient moment. Spontaneous expressions of emotion are distinct from reflexes like blushing when you are embarrassed or sweating when you are anxious. These reflexes seem entirely under the control of sub-personal systems. Spontaneous expressions of emotion on the other hand are behaviours that are up to us. Even when completely spontaneous, and so not preceded by any kind of conscious choice, they are still behaviours over which we are in control.

Like sub-intentional actions, spontaneous expressions of emotion are not actions we take for a reason: when we embrace a loved one or cry upon hearing bad news, we do not do these things because it is sensible or rational or good to do so. Such actions also do not seem to be accompanied by a desire to achieve a goal and a belief about how to achieve that goal. Of course, you *can* express an emotion in order to achieve something. For example, you might laugh at a joke not because you find it funny but in order to please the joke-teller. In this case, one could explain your laughing in terms of another activity you are engaging in, namely pleasing the joke-teller. However, examples like this are not truly spontaneous expressions of emotion. It is perhaps more accurate to describe them as emotional performances.

Furthermore, spontaneous expressions of emotion do not seem to be subject to rationalising explanations. Rosalind Hursthouse (1991) argues that spontaneous expressions of emotion cannot be explained by stating that the agent wanted to express an emotion (or vent it, or relieve it, or make it known) and believing that their actions constituted the expression of that emotion. Hursthouse correctly points out that many spontaneous expressions of emotion are simply not accompanied by a desire to express an emotion. Crying upon hearing bad news, for example, is often not something we want to do at all. Hursthouse also argues that it is wrong to suppose that the agent of a spontaneous expression of emotion possesses a belief about whether or not their behaviour constitutes an expression of the emotion they are expressing. The reason this would be wrong is because when an agent spontaneously expresses an emotion they *cannot be wrong* about whether what they are doing constitutes an expression of the emotion they are expressing. If I am crying to express my sadness, I cannot be wrong about whether my crying is an expression of sadness or not. Hence, it does not make sense to ascribe to me the belief that my crying

is an expression of sadness and to use this belief to explain why I am crying. This contrasts with actions which are subject to rationalising explanations. For example, when we explain why Carlin is adding rosemary to the sauce by stating that he wants to make the sauce taste better and believes adding rosemary will accomplish that, Carlin can be wrong about whether adding rosemary will make the sauce taste better. For this reason, it makes sense to ascribe to Carlin the belief that adding rosemary will make the sauce taste better and to use this belief to explain his action.

A third group of actions that do not display the typical characteristics of intentional actions is the actions of non-human animals. It is controversial whether non-human animals are capable of agency. We naturally speak of non-human animals doing things using the very same verbs we would use to describe some human actions: non-human animals hunt, seek shelter, raise young, climb, explore, cower, fight. However, in philosophy of action it is widely accepted that not everything an animal can be said to 'do' counts as an action of that animal. It is perfectly legitimate to speak of forgetting or falling over as things that one has done, even though forgetting and falling over are not, in any sense, actions. Reflex behaviours too can be things that we do—but they are not usually considered demonstrations of agency. There is a distinction between genuine actions, which are demonstrations of agency, and so-called 'mere behaviour': bodily movements that do not count as demonstrations of agency. It is controversial whether the bodily movements of animals count as agential or as mere behaviour.

One reason that philosophers have been reluctant to count the actions of non-human animals as demonstrations of agency, as opposed to mere behaviour, is because they have doubted that animals are capable of acting intentionally. Such arguments often rely on the assumption that animals lack the mental capacities that are prerequisites for intentional action.<sup>20</sup> For instance, it is doubtful that non-human animals are able to think of their actions as sensible or rational or good because non-human animals probably lack the ability to assess how well different courses of action could execute their intentions.

Many actions of non-human animals can be described as goal-directed. We often describe non-human animals as trying to do certain things. For example, the cat is trying to catch the mouse; the mouse is trying to hide. Furthermore, their behaviour demonstrates the kind of plasticity or flexibility we would expect if we assumed that they were acting in pursuit of a goal. If swiping towards the mouse's hiding place is unsuccessful, the cat might try waiting in ambush instead. Animals certainly seem to behave as if they were pursuing goals: they adjust their behaviour in response to changes in their environment, they change their behaviour to overcome obstacles, and they employ new tactics if their first attempts fail. However, Mele suggests that 'intentional action

---

<sup>20</sup> For example, Davidson (1982), Hacker (2007), McDowell (1996) and Stoeker (2009).

is not merely goal-directed action, but action directed in light of the agent's own goals, or desires; and desires, perhaps typically in conjunction with beliefs linking desired goals to prospective instrumental behaviour arguably constitute reasons for action' (1992b: 200). Even if animal behaviour is goal-directed, it is doubtful that non-human animals possess beliefs that link their goals to prospective instrumental behaviour.

We often successfully explain animal behaviour in terms of what the animal wants and believes. For example, the cat wants to catch the mouse; the mouse believes that under the sofa is a good hiding place. However, it is unclear that these explanations qualify as genuine rationalising explanations. Rationalising explanations, remember, explain why an agent acted as she did by telling us why, *in the agent's eyes*, what she did was a rational thing for her to do. It is not sufficient, then, that an attribution of a belief–desire pair makes the action intelligible *to us*. It is also necessary that the agent herself recognises that her desires and beliefs rationalise her actions. The agent needs to recognise that their action is desirable because it satisfies their own desire. It is at least questionable that non-human animals are able to do this.

The above examples contradict an important thesis which many supporters of the causal theory of action accept, which is that all actions are intentional under a description. The most obvious reply the causalist could make is to say that these examples are not really actions at all. The thought would be that sub-intentional action, spontaneous expression of emotion and the actions of non-human animals are not sufficiently distinct from passivity to qualify as actions at all.

Helen Steward (2009a) argues against this suggestion. She points out that it is completely natural to ascribe the production of the movements associated with sub-intentional actions to the person: 'when I fiddle with my jewellery ... it is *me* who is fiddling with it, even if I am not aware that I am doing so' (2009a: 300). Steward thus has the opposite intuition to Velleman about these cases. Sub-intentional actions would count as *sub-par actions* for Velleman, and so they would be the kinds of actions for which Velleman is happy to say that the agent is not involved. Steward thinks the agent is very much involved in sub-intentional actions, and I agree. When I tap my feet to the music, it is me who is doing so. The movement is attributable to me as a person even if I am not performing this movement for the sake of any end or even with any awareness. Steward also emphasises that the agent of a sub-intentional action is *active* in bringing about the movements; it makes sense to speak of the person moving their body in these cases. To illustrate, consider how tapping your foot to music is very different to moving your foot because a doctor has triggered your patella reflex by tapping your knee. In the former case we would comfortably say that you are moving your body, even if you are not doing it on purpose; in the latter case we would say that your foot moved but not that you moved it. Furthermore sub-intentional actions are under the agent's control: 'The fiddling seems to be something which is under my *control*, and I seem to control it in

very much the same way that I control many of the processes which constitute my intentional actions (although in the sub-intentional case, the control is not exercised in the service of an end)' (2009a: 300). The agent of a sub-intentional action seems to possess exactly the kind of control over their movements that is lacking in deviant causal chain cases.

A similar argument can be made about spontaneous expressions of emotion and the actions of non-human animals. Spontaneous expressions of emotion are attributable to the person: no other agent, or sub-personal system, is acting through them. It is natural to speak of the person moving their body in cases of spontaneous expressions of emotion. And, even when completely spontaneous, expressions of emotion are behaviours over which we are in control. Another consideration that speaks in favour of counting spontaneous expressions of emotion as examples of agency is the fact that so much of our behaviour is emotionally driven. We might like to think that most of our actions are fully intentional, that most of our actions are done in pursuit of a goal, that we decide to do most of what we do, but I think that is wishful thinking. A great deal of what we do is done as an expression of emotion. In many situations, there simply isn't time to think about what to do before taking action. Choices are made, directions given, words spoken before any beliefs about the situations that called for those choices, directions or words have been formed. A lot of the time, we act spontaneously, using our feelings about a situation to guide us rather than our thoughts.

Similarly, at least some non-human animals seem capable of controlling their bodies in exactly the way that subjects in deviant causal chain cases cannot. As mentioned, many non-human animals display the kind of flexibility in their behaviour we would expect if they were acting in pursuit of a goal. They adjust their behaviour, they overcome obstacles, they try again if they don't succeed. This seems to imply that animals direct their own movements as opposed to passively undergoing changes in response to events occurring in their environment or inside their bodies. We also often speak of animals as moving their bodies and attribute their movement to them. As Steward argues elsewhere:

It is most unnatural to suppose that the cockerel was caused to make its journey across the yard by anything like a mere reflex or a simple stimulus-response mechanism. For although we obviously have to recognise the huge importance of instinct in the lives of animals, instincts which prescribe for a given animal a range of basic activities from which it is certainly not free to forbear, I think we allow to the animal—and this is crucial, in my view, for the concept of agency—a certain freedom and control over the precise movements by means of which it satisfies those instinctual needs and desires. (2009b: 225)

Steward further defends her intuition that sub-intentional actions are genuine exercises of agential power by arguing against a line of thought that would

pull someone in the opposite direction. Her argument here could apply equally well to spontaneous expressions of emotion and animal action. Someone who wanted to discount these examples as actions might think that ‘unless there is some reason to suppose that a movement is in some sense the product of something mental, there can be no reason to think it should be associated in any special way with the self, with the agent ... Unless my mind is somehow involved, the thought goes, *I* could not be involved either’ (2009a: 303). Steward argues that this thought stems from two prejudices.

The first is the Cartesian assumption that, if a person can control her own body, then the thing doing the controlling in that case must be the person’s mind. Steward suggests that we think of some animals as being in possession of their bodies: some animals have bodies that they can to some extent control: ‘We think and speak of animals—especially human ones—as *possessed of* their bodies, and to a certain extent, as controllers of them’ (2009a: 303). However, this innocuous thought does not entail that, when an animal controls its movements, its mind controls its movements. Steward acknowledges that we typically attribute body-possession and mindedness together. It is an important truth that properties like having a mind, having thoughts and being conscious seem conceptually connected to properties like being the kind of creature that has a body it can control. However, Steward insists that the existence of this important connection does not entail that every time an animal controls its body this must be a case where the animal’s *mind* controls its body. What is suggested by Steward’s argument is that causal theories of action ‘over-mentalise’ agency by assuming that being able to control one’s body entails the existence of a mind doing the controlling.

The second prejudice concerns the nature of causation. The assumptions about causation that Steward thinks prevent accurate appraisal of sub-intentional action, spontaneous expressions of emotion and animal actions are precisely the assumptions which constitute relationalism. Relationalism says that causal reality is nothing more than a chain of causally related events. This means that the causal truths about agency must be truths concerning causation of and by certain events; therefore, any distinction crucial to our conception of agency must be a distinction between different types of event causation. If one is committed to relationalism, then the idea of an animal controlling its movements *must* be reducible to a statement about an event occurring within the animal that produces the effect. This is why the distinction between intentional actions and other events becomes very important, because there is plausibility to the idea that mental causation is key to understanding *this* distinction. For Steward, this constitutes a *prejudice* because it forces us to think that, if non-intentional actions are actions at all, then they must have a mental cause instead of taking them at face value: genuine exercises of agential power that do not have mental causes.

Sub-intentional action, spontaneous expressions of emotion and the actions of non-human animals are not counterexamples to the causal theory of action.

They do not disprove the causal theory of action, as that theory is only intended to be a theory of *intentional* actions—it is not required to say anything about actions which are not intentional. However, any theory of intentional action should recognise the *continuity* between intentional action and other forms of agency. There is continuity between the kind of control demonstrated in non-intentional action and the kind of control demonstrated in intentional action. It would be wrong, I think, to say that these are examples of a completely different kind of control. Rather, intentional action is a *development* of the kind of agential control demonstrated in sub-intentional action, spontaneous expressions of emotion and animal action—it is the same fundamental phenomenon but extended or enhanced. The physicalist/event-causal picture of agency is poorly equipped to recognise this continuity as it ties the agency concept so closely to intentionality and mental causation.

## 4.2 Conclusion

I have presented three criticisms of the physicalist/event-causal picture of agency: it fails to solve the problem of deviant causal chain cases; it fails to account for refrainment; and it fails to account for the unity between intentional action and non-intentional agency. I now need to explain how these three failures connect to the disappearing agent problem.

The causal theory of intentional action aims to understand intentional action via a single divide: between event-causal sequences that involve intentional states and those which do not. However, the boundary between agential and non-agential does not map onto this divide. The two distinctions cut across each other. Sometimes a certain kind of mental causation is what stops an example counting as an instance of agency (deviant causal chain cases); our agency concept extends to cases where agents remain passive and so no bodily movement is caused to happen (refrainment); and our concept of agency extends to cases where there is no mental cause of a bodily movement (non-intentional action). 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.

Common to all the diverse examples of agency described above is *the involvement of the agent*. In both Velleman's action par excellence, where an agent 'moves his limbs in execution of his intention,' and sub-intentional action, where the agent is barely aware that they are moving their limbs and intentional states play no causal role, the agent is in control of their body—their bodily movements are up to them. When an agent refrains from doing something, the agent still retains some control over the situation in virtue of not exercising a power to act. And even if we cannot confidently say that animals act for reasons, or ascribe to them the propositional attitudes typically associated with intentional agency, they still seem to have control over the movement of their bodies.

The best explanation for why one cannot provide a comprehensive account of agency if one abides by the assumptions of the physicalist/event-causal picture of agency is because this picture leaves the agent out. If one assumes that causal reality is nothing more than a chain of causally related events, and therefore that the causal truths about agency are truths concerning causation of and by certain events, then any distinction crucial to our conception of agency must be a distinction between causal relations involving a mental relatum and causal relations that do not involve a mental relatum. However, this assumption leaves us unable to resolve the three issues described above. The distinction between agency and non-agency does not map onto a distinction between causation involving mental causes and causation not involving mental causes.

To adequately understand agency we need a metaphysical framework that allows us to see how the causality of action might be something that casts the agent herself as a causal player, rather than merely the setting for events to cause other events. The physicalist/event-causal picture of agency is unsatisfactory because our general concept of agency is fundamentally at odds with a view of the world that assumes that causal reality is nothing but a chain of causally related events.

### References<sup>21</sup>

- Alvarez, M 2013 Agency and two-way powers. *Proceedings of the Aristotelian Society*, 113(1pt1): 101–121. DOI: <https://doi.org/10.1111/pash.2013.113.issue-1pt1>
- 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.
- Clarke, R 2014 *Omissions: Agency, metaphysics, and responsibility*. New York: Oxford University Press.
- Clarke, R 2017 Free will, agent causation, and ‘disappearing agents.’ *Noûs*, 53(1): 76–96. DOI: <https://doi.org/10.1111/nous.12206>

---

<sup>21</sup> Author note: some references to Davidson are formatted (1963/2001a). This indicates the initial date of publication of the paper (in this case 1963) but references the paper as it appears in the 2001a collection of his essays, with the page numbers relating to that volume. Similarly, some references to Davidson are formatted (1997/2001b) which indicates the initial date of publication (in this case 1997) but references the paper as it appears in the 2001b collection of Davidson’s essays, with the page numbers relating to that volume.

- Davidson, D 1971 Agency. In: Binkley, R, Bronaugh, R and Marras, A *Agent, action, and reason*. Toronto: University of Toronto Press. pp. 1–37. Reprinted in Davidson 2001a pp. 43–62.
- Davidson, D 1973 Freedom to act. In: Honderich, T *Essays on freedom of action*. New York: Routledge and Kegan Paul. pp. 137–156. Reprinted in Davidson 2001a pp. 63–82.
- Davidson, D 1982 Rational animals. *Dialectica*, 36(4): 317–328. Reprinted in Davidson 2001b pp. 95–106.
- Davidson, D 2001a *Essays on actions and events*. 2nd ed. Oxford: Clarendon Press.
- Davidson, D 2001b *Subjective, intersubjective, objective*. Oxford: Clarendon Press.
- De Swart, H 1996 Quantification over time. In: van der Does, J and van Eijk, J *Quantifiers, logic, and language*. Cambridge: Cambridge University Press.
- 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.
- Frankfurt, H 1969 Alternate possibilities and moral responsibility. *Journal of Philosophy*, 66(23): 829–839. DOI: <https://doi.org/10.2307/2023833>
- Hacker, P 2007 *Human nature*. Oxford: Blackwell.
- Higginbotham, J 2000 On events in linguistic semantics. In: Higginbotham, J, Pianesi F and Varzi, A *Speaking of events*. New York: Oxford University Press, pp. 49–81.
- Hornsby, J 2004 Agency and alienation. In: Macarthur, D and De Caro, M *Naturalism in question*. Cambridge, MA: Harvard University Press. pp. 173–187.
- Hursthouse, R 1991 Arational actions. *Journal of Philosophy*, 88(2): 57–68. DOI: <https://doi.org/10.2307/2026906>
- Hyman, J 2015 *Action, knowledge, and will*. New York: Oxford University Press.
- Lewis, D K 1986 *Philosophical papers*. New York: Oxford University Press.
- Lewis, D K 1994 Humean supervenience debugged. *Mind*, 103(412): 473–490. DOI: <https://doi.org/10.1093/mind/103.412.473>
- McDonnell, N 2015 The deviance in deviant causal chains. *Thought: A Journal of Philosophy*, 4(2): 162–170. DOI: <https://doi.org/10.1002/tht3.169>
- McDowell, J 1996 *Mind and world*. Cambridge, MA: Harvard University Press.
- Melden, A I 1961 *Free action: Studies in philosophical psychology*. London: Routledge & Kegan Paul.
- Mele, A 1992a *Springs of action: Understanding intentional behavior*. New York: Oxford University Press.
- Mele, A 1992b Recent work on intentional action. *American Philosophical Quarterly*, 29(3): 199–217.
- Mele, A 2000 Goal-directed action: Teleological explanations, causal theories, and deviance. *Noûs*, 34(14): 279–300. DOI: <https://doi.org/10.1111/0029-4624.34.s14.15>

- Mele, A 2003 *Motivation and agency*. Oxford: Oxford University Press.
- Mele, A 2005 Action. In: Jackson, F and Smith, M *The Oxford handbook of contemporary philosophy*. Oxford: Oxford University Press. pp. 78–88.
- Nagel, T 1986 *The view from nowhere*. New York: Oxford University Press.
- O’Shaughnessy, B 1980 *The will*. Cambridge: Cambridge University Press.
- Paul, S 2020 *Philosophy of action: A contemporary introduction*. London: Routledge.
- Peacocke, C 1979 Deviant causal chains. *Midwest Studies in Philosophy*, 4(1): 123–155. DOI: <https://doi.org/10.1111/j.1475-4975.1979.tb00375.x>
- Pereboom, D 2014 *Free will, agency, and meaning in life*. New York: Oxford University Press.
- Schaffer, J 2007 Causation and laws of nature: Reductionism. In: Sider, T, Hawthorn, J and Zimmerman, D W *Contemporary debates in metaphysics*. Malden: Blackwell, pp. 82–107.
- Sehon, S 1997 Deviant causal chains and the irreducibility of teleological explanation. *Pacific Philosophical Quarterly*, 78(2): 195–213. DOI: <https://doi.org/10.1111/1468-0114.00035>
- Shepherd, J 2021 *The shape of agency: Control, action, skill, knowledge*. Oxford: Oxford University Press.
- Smith, M 2010 The standard story of action: An exchange 1. In: Buckareff, A A and Aguilar, J H *Actions: New perspectives on the causal theory of action*. Cambridge, MA: MIT Press. pp. 45–56.
- Smith, M 2012 Four objections to the standard story of action (and four replies). *Philosophical Issues*, 22(1): 387–401. DOI: <https://doi.org/10.1111/j.1533-6077.2012.00236.x>
- Steward, H 2009a Sub-intentional actions and the over-mentalization of agency. In: Sandis, C *New essays on the explanation of action*. Basingstoke: Palgrave Macmillan.
- Steward, H 2009b Animal agency. *Inquiry*, 52(3): 217–231. DOI: <https://doi.org/10.1080/00201740902917119>
- Stoecker, R 2009 Why animals can’t act, *Inquiry*, 52(3): 255–271. DOI: <https://doi.org/10.1080/00201740902917135>
- Velleman, D J 1992 What happens when someone acts? *Mind*, 101(403): 461–481. DOI: <https://doi.org/10.7591/9781501721564-008>
- Vermazen, B 1985 Negative acts. In: Vermazen B and Hintikka, M B *Essays on Davidson: Actions and events*. Oxford: Clarendon Press. pp. 93–104.
- Yablo, S 1992 Mental causation. *Philosophical Review*, 101(2): 245–280. DOI: <https://doi.org/10.2307/2185535>
- Zagzebski, L 1994 The inescapability of Gettier problems. *Philosophical Quarterly*, 44(174): 65–73. DOI: <https://doi.org/10.2307/2220147>