

CHAPTER 5

Agent Causation

In the previous chapter, I argued that we should try to break out of the physicalist triad since it provides an inadequate account of agency. The main failing of the physicalist/event-causal account of agency entailed by the physicalist triad is that it cannot provide a comprehensive account of agency—one that solves the problem of deviant causal chains, explains why refrainment counts as intentional action and accounts for the unity between intentional action and non-intentional action. The physicalist/event-causal account of agency is unable to deliver a comprehensive account because it leaves out the agent. This is the disappearing agent objection, and, although the objection is often misunderstood (see Section 4.1), I believe it is the most powerful objection against a physicalist/event-causal account of agency. The point of this objection 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. Thus, what is needed to adequately understand agency is a richer theory of causation, one 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.

Philosophers working within the field of philosophy of action and on the problem of free will have offered theories of what agency is which attempt to avoid the disappearing agent objection. Many of these accounts appeal to the notion of *agent causation*. According to this general type of view, agency is a kind of causation where the agent, who is taken to be a substance not an event, exercises causal power and this exercise of causal power cannot be reduced to causation by an event involving the agent. So, for example, what makes my action of typing this sentence a demonstration of agency is that *I* am causing letters to appear on my computer screen, where this *causing* of mine cannot be understood as the causation of one event by another (e.g. the causation of finger movements by a decision to type)—it is its own special type of causation.

How to cite this book chapter:

White, Andrea. 2024. *Understanding Mental Causation*. Pp. 97–114. York: White Rose University Press. DOI: <https://doi.org/10.22599/White.f>. License: CC BY-NC 4.0

Appealing to agent causation to explain what agency is represents a departure from the standard relationalist view of causation, which takes all causation everywhere to be a relation between events or states. However, as we shall see, many agent-causationist accounts of agency accept some aspects of relationalism about causation and face significant issues as a result. In what follows, I will critically examine some agent-causal accounts of agency and argue that the chief failing of these theories is that they do not go far enough when it comes to rejecting relationalism about causation.

5.1 Traditional agent-causationism

I will first examine what I call ‘traditional agent-causationism’. This title covers a family of theories which maintain that irreducible agent causation is required to adequately explain aspects of specifically human action. Traditional agent-causationists maintain that human agency is causation by an agent, who is taken to be a substance. According to this view, agent causation is a form of causation that cannot be identified with, or realised by, a causal relation between events or states. As such, causation by an agent cannot be analysed in terms of causation by any event involving the agent—causation by an agent is, in this sense, ontologically fundamental. An important tenet of traditional agent-causationism is that it is specifically human actions that must be understood as examples of irreducible agent causation. Traditional agent-causationists accept that *most* causation in the world, including interactions between non-human animals and inanimate objects, is nothing over and above causation of one event by another. It is only in the case of things done freely by human agents that there is something extra—causation by a *substance*.

Traditional agent-causationism is usually motivated by considerations to do with free will. Recall Pereboom’s (2014) argument for understanding free will in terms of agent causation. For Pereboom, it is specifically *free* action that must involve irreducible agent causation because free agents need to, themselves, be the *determiners* or *settlors* of their actions. Roderick Chisholm also argues that agent causation is essential for an adequate treatment of *free will* (1976: 58–59). Richard Taylor similarly rests his case for an agent-causation-based account of agency on the idea that agents must be the ‘initiators’ or ‘originators’ of their actions (Taylor 1966: 112) and argues that this sense of ‘initiation’ or ‘origination’ is lacking in cases where inanimate objects cause things to happen. Taylor thus commits himself to the view that inanimate objects are never agents: ‘a man is sometimes an agent who originates a change, and is not, like a match, merely a passive object that undergoes change in response to other changes’ (1966: 122). Taylor denies that a match can be an agent because a match cannot ‘wreak changes in itself’: what a match does is always a response to the circumstances it is in and what’s acting upon it. A person, in contrast, ‘can bring about such a change as a motion of his arm quite by himself’ (1966: 122). For this

reason, human agency must be understood in terms of an irreducible form of substance causation, but there is no similar demand to understand causation by inanimate objects in terms of irreducible substance causation.

Timothy O'Connor (2000; 2009) also argues that substance causation is a form of causation uniquely exercised by persons. O'Connor argues that 'an adequate account of freedom requires, in my judgement, a notion of a distinctive variety of causal power, one which tradition dubs "agent-causal power"' (2009: 230). In essence, O'Connor's view is that in order to make free choices about how we act, our actions need to be 'up to us'. Our actions are not up to us if prior events deterministically cause them, or so the thought goes. However, their being up to us also cannot consist in our actions being the non-deterministic causal consequence of certain events because, as O'Connor (2009: 231) puts it, 'looser connectivity in the flow of events' cannot constitute any kind of personal control over what happens. As O'Connor writes:

[I]f I am faced with a choice between selfish and generous courses of action, each of which has some significant chance of being chosen, it would seem to be a matter of luck, good or bad, whichever way I choose, since I have no means directly to settle which of the indeterministic propensities gets manifested. (2009: 231)

The solution O'Connor endorses is to endow agents with a special causal power to bring about events, a power they must exercise if they are to act freely. In essence, the motivation for traditional agent-causationism is that agents themselves—and no events involving the agent—must cause their actions, otherwise free action is metaphysically impossible.

There are three key points to note about the metaphysics traditional agent-causationists think is required for agents to act freely. First, when an agent causes an event, the agent causes that event directly, which is to say that no event involving the agent or circumstances about the agent cause the event; in fact the event that the agent causes has no cause other than the agent herself. Second, agents cause their own actions. Third, there is no demand to understand causation by inanimate objects in terms of irreducible substance causation; in this way, human agency is metaphysically exceptional.

According to traditional agent-causationism, the event of my raising my arm, which is my causing my arm to rise, is an action because it is an event that I, *qua* substance, caused to happen. However, there is a well-known problem with this view. If my action is an event of which I am the cause, then we can ask of the causing of my action whether *this* is an action of mine or not. If it is, then, on the agent-causationist theory, it is also an event of which I am the cause, but now we seem to have opened an infinite regress: is the causing of my causing of my action another action? However, if we deny that the causing of my action is an action, then it seems we have two sorts of 'causings', some of which are actions and some of which are not. For example, my causing my arm to rise is

an action, but the causing of my causing my arm to rise is not—what makes this difference? It is unclear what the agent-causationist can or should say.²²

Another important objection to traditional agent-causationism is that the account of agent causation I have just summarised is metaphysically unintelligible. Agent causation is supposed to be a special kind of causation distinct from and not reducible to a causal relation between events. The question is: what exactly is this special type of causation? What is it for an agent to ‘directly cause an event’ if this cannot be reduced to causation by an event involving the agent? Sceptics of agent causation argue that we have no independent understanding of what agent causation is. For example, von Wright (1971: 192) argues that the only way to make sense of agent causation is to see it as a synonym for human agency. This is especially the case if irreducible substance causation only exists in cases where a human being is acting freely. Limiting irreducible substance causation to exercises of free will makes substance causation seem like something discontinuous with the non-human world. Substance causation is made to seem like something additional to the world’s normal causal functioning, which appears only when human beings act freely. Erasmus Mayr argues that:

[R]estricting agent-causal activity to human agents (among the objects in the world) tends to make agent causation appear either as some unnatural extra force with which human beings are endowed and which can only be compared to divine causation—a comparison which is unlikely to improve our understanding of the notion—or as simply another name for the phenomenon we want to understand: human agency. (2011: 143)

The solution, which has been proposed by Mayr (2011) and Helen Steward (2012), is to insist that irreducible substance causation is ubiquitous. Non-human animals and inanimate objects cause things to happen in the same sense in which human beings cause things to happen when they act. In other words, substances causing things to happen is a general feature of the world, and human agency is just a special example of it. In no case can causation by a substance be reduced to causation by some event involving that substance. This may be because, in fact, all causation is fundamentally substance causation (Lowe 2008) or because causation is a diverse phenomenon and entities of many different categories—substances, events, facts, properties—can cause things, and each of these types of causation is fundamental (Hyman 2015; Mayr 2011; Steward 2012).

Even though the metaphysics proposed by traditional agent-causationists is ultimately unsuccessful, there are some aspects of traditional agent-causationism that I think are correct. What’s right about traditional agent-causationism

²² See Alvarez and Hyman (1998), Davidson (1971) and Hornsby (1980) for discussions of this problem.

is that the actions of humans (and I believe many non-human animals) are importantly metaphysically different from causation by inanimate objects. I think it is right to suppose, as O'Connor does, that human (and animal) agency must be a causal power of a special kind. The metaphysical exceptionalism of human (and animal) action is borne out in experimental philosophy. John Turri (2018) summarises findings from experiments that seem to suggest people think that 'human agency fits broadly within the causal order while still being exceptional in some respects' and more specifically that 'people believe human actions are caused by a variety of factors, including psychological, neurological, and social events' (2018: 402) and in that respect are part of the same causal order as everything else, but that humans and animals (but not computers or plants) were always capable of acting otherwise even if 'everything in the causal history of the physical world' rendered a certain outcome 100% probable (2018: 407). These findings do not tell us how to understand human agency, or how to spell out what is distinctive about it, but they demonstrate the pervasiveness of the intuition that human (and animal) agency is exceptional in some way.

I also think that traditional agent-causationists are right to seek to explain the metaphysical exceptionalism of human (and animal) agency by reference to the idea of things being up to the agent. I agree that it seems to be an essential part of our concept of agency that acting must involve a minimal kind of autonomy.

An essential characteristic of agency is that, when an agent acts, some of what goes on with the agent is up to the agent. One way to elucidate the idea of things being 'up to' the agent is to make use of Aristotle's distinction between self-movement and moved-movement. Humans and animals can move themselves; they do not need to be pushed or prodded or pulled by something else in order to move. Inanimate objects, on the other hand, can move, but they must be 'moved to move' by some other thing.

To help illustrate the distinction, consider the following examples. When a stone is thrown at a window with sufficient force, there is no sense in which it is *up to the stone* whether or not it breaks the window. If the conditions are right, i.e. the stone is heavy enough and the glass is thin enough, the stone will break the window (provided nothing comes along and interferes, e.g. no-one snatches the stone out of the air before it hits the window). The stone may well be the thing that is breaking the window—in this way the stone is a 'mover' (or, more precisely, a 'breaker')—but it was 'moved' to do so; that is, the stone was directed to break the window by some other thing (whatever threw it). Now consider the child who threw the stone. Ordinarily, when a child throws a stone the child moves his own body to move the stone. Even if the child was acting out of such intense emotion that we would not want to say the action was free, or intended, or chosen, no other thing is moving the child's arm for him. In this sense, the child is moving himself.

Robots that mimic human movements, such as Honda's ASIMO, are also moved-movers and not self-movers. This might seem counterintuitive because, unlike a stone, ASIMO can move around and perform various tasks without

another substance intervening. However, ASIMO's movements are strictly governed by his construction and programming. To illustrate: ASIMO has two cameras, a laser sensor, an infrared sensor and an ultrasound sensor. When information recorded by these sensors conflicts with information in ASIMO's pre-loaded map of navigable paths (e.g. by signalling that there is an obstacle in one of these paths), ASIMO cannot but move around the obstacle (American Honda Motor Co. Ltd. Public Relations Division 2007). ASIMO is moved to move around the obstacle by his component parts. It is not up to ASIMO what goes on with his legs. It is a necessary condition on our movements being up to us, and hence being genuine demonstrations of our agency, that we are not moved to move by our component parts.

Although I agree with traditional agent-causationists that there is an important metaphysical difference between human (and animal) action and causation by inanimate objects and that the metaphysical exceptionalism of human and animal agency has something to do with things being up to the agent in the former case but not the latter, I disagree that this metaphysical difference should be explained as the difference between two kinds of causation with different *ontologies*. I agree that human action is distinctive but I disagree that what makes human action distinctive is that it involves irreducible substance causation, whereas all other causation in the world is nothing over and above event causation. This is the wrong way to explain what makes human action exceptional. At best it makes human action seem like something unnatural. The capacity for self-movement is made to seem like a god-like capacity to directly interfere with event-causal chains. At worst it introduces a form of causation, causation of an event by an agent, which can only be understood as a synonym for human agency.

The reason writers like Pereboom, Chisholm, Taylor and O'Connor have gone wrong is, I think, because they have rejected relationalism about causation only in part. They accept the standard relationalist picture of causality with respect to animals and inanimate objects causing things to happen but reject it in the case of human agency. So, for example, when a stone breaks a window, the *real* cause is the event of the stone's being thrown towards the window, not the stone, but when a person breaks a window, the real cause is the person, not any event. However, this piecemeal departure from the relationalist picture is not justified.

One line of thought that might lead one to think that there must be substance causation in the case of human agency, but not in cases where inanimate objects cause things to happen, is as follows. Because inanimate objects are moved-movers, they are passive, which is to say they never cause change; they only suffer change. Therefore, the *real* cause in cases where an inanimate object makes something happen must be an event. However, this reasoning is fallacious. It is a fallacy to conflate moved-movement with passivity. Passivity is the manifestation of a passive power, or a liability, i.e. a power to undergo or suffer change. It contrasts with activity, which is the exercise of an active power,

i.e. a power to wreak change. Active powers are powers to change, and passive powers are powers to be changed. As John Hyman points out, the difference between agent and patient is not a difference between two different kinds of substance; it is rather a difference between two different roles substances can adopt (2015: 35). It is also possible for one and the same substance to be both agent and patient at the same time. For example, as Hyman notes, a victim of suicide is both agent and patient. Moved-movers when they cause change are both active and passive: active because they are causing a change but passive as well because their causing that change is dependent on another substance acting upon them.

A similar consideration that might lead one to think that there must be substance causation in the case of human agency, but not in cases where inanimate objects cause things to happen is discussed by Steward (2012). Steward considers the suggestion that, in cases where an inanimate object brings about an event, 'it is usually true that the object would not have caused the effect in question had it not been involved in some relevant event' (2012: 208). For example, the stone would not have broken the window had the child not thrown it. It would not have broken the window had it remained on the ground. From this, we may conclude that the event the stone is involved in is the real cause of the window-breaking. However, Steward argues that this reasoning is also fallacious. It depends on confusing causation and causal explanation. In order to adequately explain how the window came to be broken, we need to say something about how the stone came to break the window. It is rarely sufficient to answer the question 'why is the window broken?' by stating 'because of the stone.' However, as Steward points out, the fact that an adequate explanation requires reference to an event does not allow us to conclude that the stone 'does no causal work' (2012: 209). In Chapter 9, I will offer a positive account of how I think the crucial contrast between self-movement and moved-movement should be understood.

5.2 Actions-as-causings

The second agent-causation-based account of agency I shall consider does not contend that agents cause their own actions. According to this alternative agent-causation-based account, an agent's action *is* her causing of something; it is not what is caused. As Maria Alvarez and John Hyman put it, 'an action is a causing of an event by an agent' (1998: 224). I shall call this kind of theory the 'actions-as-causings' view. According to the actions-as-causings view, agency consists in an agent coming to stand in a causal relation to an event, or sometimes a state of affairs. However, what the agent causes is not her own action, instead the agent causes an event 'intrinsic' to the agent's action, an event that Alvarez and Hyman (1998: 233) call the 'result' of the action. The result of an action is not a causal consequence of the action; the relationship between

an action and its result is much tighter than that. For example, the result of an action of answering the phone is the event of the phone being answered. The action is what the agent does, and the result of the action is what must happen if the action is actually performed. Often, in the case of human action at least, the ‘result’ of an action is a bodily movement. For example, my action of raising my arm consists in my causing the rising of my arm. The rising of my arm is the result of my action and the event intrinsic to my action. *I* am the cause of my arm-rising, and my so being the cause of my arm-rising is what my action consists in.

The actions-as-causings view is most explicitly endorsed by Alvarez and Hyman (1998). However, Mayr also argues that human agency is an instance of substance causation (2011: 219), where substance causation should be understood in terms of a causal relation obtaining between a substance exercising an active power and the effect produced when the substance exercises active power: ‘when such an “active power” is exercised, the cause of the resulting event is the substance which possess the power itself’ (2011: 145–146). Similarly, E. J. Lowe describes agent causation as a species of causation ‘in which the cause of some event or state of affairs is not (or not only) some other event or state of affairs, but is, rather, an agent of some kind’ (2008: 121).

The crucial feature of the actions-as-causings view is that agency is described in terms of a causal relation, albeit one that obtains between an agent and an event or state of affairs. According to the actions-as-causings view, to properly understand agency we need to recognise that agents, *qua* substances, can be causes. The actions-as-causings view thus departs from standard relationalism insofar as it allows that substances can be relata of the cause–effect relation, not just events. However, substance causation is still described in relational terms. The action-as-causings view still accepts that causation is a relation between cause and effect; it just allows that substances—as opposed to only events—can be causes.²³

My objection to the actions-as-causings view is that it entails two counterintuitive claims. First, the actions-as-causings view entails that one’s actions are never identical to the bodily movements one’s body makes when one acts. So, for example, my raising my arm cannot be identical with my arm’s rising. Alec Hinshelwood calls this claim ‘the separation thesis’ (2013: 626). The second

²³ This kind of view is also endorsed by Harré and Madden (1975), who defend an account of causation as powerful particulars, which are substances, producing effects. For example, when a rock breaks a window, it comes to stand in a production relation to a window-breaking event. Thomas Reid also thought that causation was the production of change by the exertion of power and ‘that which produces a change by the exertion of its power we call the *cause* of that change; and the change produced, the *effect* of that cause’ (1788: 12–13).

counterintuitive claim the actions-as-causings view entails is that actions are not events.

For proponents of the actions-as-causings view, the separation thesis and the idea that actions are not events should not be seen as reasons to reject the actions-as-causings view. Instead they should be viewed as interesting, and inevitable, consequences of accepting that agency ought to be understood in terms of agent causation. However, this is incorrect. The separation thesis, and the idea that actions are not events, are not direct consequences of accepting that agency ought to be understood in terms of agent causation. Instead, these views are entailed specifically by the relational interpretation of agent causation endorsed by the actions-as-causings view.

5.2.1 *Two counterintuitive claims*

Alvarez and Hyman (1998) explicitly argue that actions are never identical to the movements one's body makes when one acts. Here is their argument:

Davidson is one philosopher who claims that, in some cases, 'my raising my arm and my arm rising are one and the same event'. But my raising my arm is my causing my arm to rise. Hence, if my raising my arm is an event, it is the same event as my causing my arm to rise. And hence, if my raising my arm and my arm's rising are one and the same event, then my causing my arm to rise and my arm's rising are one and the same event. But it cannot be plausible that causing an event to occur is not merely an event itself, but the very same event as the event caused. (1998: 229)

Spelt out, the argument runs as follows:

Assume for *reductio*:

1. My raising my arm is one and the same event as my arm's rising.

Now assume the very plausible:

2. My raising my arm is my causing my arm to rise.

And:

3. If my raising my arm is one and the same event as my arm's rising, then my causing my arm to rise is my arm's rising.

Together these premises entail:

4. My causing my arm to rise is my arm's rising.

A conclusion that, when generalised, is revealed to be absurd:

5. My causing an event is the event caused.

In response to this argument, Alvarez and Hyman, and many writers sympathetic to agent-causation-based theories of action, have rejected the thesis that one's action is identical with the bodily movements one's body makes when one acts (the separation thesis).

To explain how the separation thesis is compatible with the plausible claim that many actions are bodily movements, Alvarez and Hyman (1998) make use of an ambiguity associated with the word 'movement' noted by Jennifer Hornsby (1980). Many verbs can be transitive (i.e. used with a grammatical object) or intransitive (i.e. used without a grammatical object). The verb 'move' is also ergative, which means that it can be transitive or intransitive and that the direct object of the verb when transitive becomes the subject of the verb when intransitive. For example, 'move' is transitive in the sentence "I moved my arm" but intransitive in "My arm moved", and the object of the transitive 'move' is the subject of the intransitive 'move'. This feature of the verb 'move' renders the nominalisation of 'move', 'movement', ambiguous. When we speak of, for example, my arm movement, there are two movements we might be talking about. There is one that corresponds to the transitive use of move, as in "I moved my arm", which can be otherwise picked out by the expression 'my moving of my arm', and the one that corresponds to the intransitive use of move, as in "My arm moved", which can be otherwise picked out by the expression 'the motion of my arm'. To help keep the two senses of 'movement' separate I will follow Hornsby's notation and use 'movement_T' for the first sense, and 'movement_I' for the second sense. Alvarez and Hyman (1998) hold that many actions are bodily movements_T, which they claim are *causings* of bodily movements_I, and hence cannot be identical with bodily movements_I.

Alvarez and Hyman (1998) also argue that actions, i.e. *causings* of bodily movements_I, are not events of any kind. To establish this conclusion, Alvarez and Hyman assume that there are only two possible sorts that event actions could be:

1. bodily movements_I; or
2. events that are causes of bodily movements_I.

Alvarez and Hyman take the first possibility to have been ruled out already by the argument outlined above. To show that bodily movements_T are not events that cause bodily movements_I, Alvarez and Hyman argue as follows:

[I]f bodily movements_T are events which cause bodily movements_I, then either bodily movements_T are events, perhaps neural events, which occur inside the agent's body, as for example Hornsby maintains in

her book *Actions*, or they are events of another sort, which do not—presumably events which have no location at all, if there are such events. The first alternative implies that bodily movements_p, unlike their effects, are not normally perceptible without a special apparatus. The second implies that bodily movements_i are caused both by neural events and by events of another sort, and therefore raises the difficult question of how these two sorts of events are related. It also implies that bodily movements_i can never be perceived, whatever sort of apparatus we are equipped with. But we can and do see people and animals moving their limbs without making use of any sort of apparatus; and seeing a person or an animal moving its limbs is seeing a bodily movement_i. Hence neither alternative is tenable; and it follows that bodily movements_i are not events which cause bodily movements_p. (1998: 229–230)

I agree that the first option Alvarez and Hyman consider here, that all actions are events that take place inside the agent, is not very plausible. Common sense suggests that many actions are public, and actions that involve moving one's body are paradigm examples of actions that other people can see without any special equipment. The second option Alvarez and Hyman consider is not as obviously implausible, partly because the option they suggest is itself difficult to understand. Explained with an example, the suggestion is that my action of raising my arm—which is assumed to be my causing of my arm's rising—is an event that causes my arm's rising but is not identical with any neural event or muscular event or indeed any of the events that occur in the vicinity of my arm's rising that are causally linked to my arm's rising. Instead it is an event that causes my arm's rising but is not located anywhere in particular. Put this way, the suggestion is very strange and Alvarez and Hyman are right to reject it.

If Alvarez and Hyman's argument succeeds, then bodily movements_i are not events, so the causing of an event by an agent is some *other* sort of entity. The actions-as-causings theory of agency thus seems to involve ontological commitment to a novel kind of entity, which is the coming-to-obtain of a causal relation between an agent and an event. To give these novel entities a name, let's call them 'causings'.

Alvarez and Hyman (1998) are not the only philosophers who argue that actions are not events. This idea has quite a long history. Kent Bach (1980) argues that actions are not events because they are the obtaining of a causal relation between an agent and an event (see also von Wright 1962 and Chisholm 1964). There is some intuitive plausibility to the idea that actions are not events because actions can be said to be things people do and you cannot 'do' an event—an event is something that happens. However, this intuition is not robust enough to support a metaphysical conclusion because the word 'action' can be used in many different ways. Often the word is used to name activities people engage in—things people do—and in that sense does not seem to refer to a set of events. For example, "She took decisive action" probably refers to

the deeds the agent undertook—the things she did—and not the events that happened. However, there are many other uses of the word where it is more plausible to assume one is talking about things that happen, i.e. events. For example, “The action surprised her” could plausibly be interpreted as referring to something that happened. Similarly, “His action triggered a revolt” also seems to reference an event.

My own view is that the idea that actions are events is, to borrow an expression from Hornsby (2004), an *innocent* one. There is nothing majorly wrong with the idea that actions are events. Although the claim that actions are events is a key claim of event-causal theories of action, it is not the claim that does the most damage to our understanding of agency. Event-causal theories of action fail to adequately explain agency because they assume causal reality is nothing but a chain of causally related events and hence that what it is to act reduces to causal relations between events. Thus, the claim that does the most damage is not that actions are events but that what makes something an action is a question of what causal relations it is involved in. The best account of agency would be one that allows that *sometimes* when we talk about actions we are talking about events because that is what our language seems to imply.

I also think that the idea that one’s actions are, at least sometimes, identical to the bodily movements one’s body makes when one acts is similarly innocent. There is nothing majorly wrong with the idea that my raising my arm and my arm’s rising are one occurrence. Indeed, I find the separation thesis counterintuitive for two reasons.

My first reason comes from an argument against the separation thesis made by Hinshelwood (2013). Hinshelwood argues that the separation thesis generates two epistemological issues, the first of which seems to me the most pressing. Hinshelwood begins his argument by pointing out that ‘we can perceive what someone is doing simply by *seeing her doing it*’ (2013: 628). In other words, actions are direct objects of perception—we can literally and directly see actions. For example, when someone raises their arm we do not see something else that serves as visual evidence of their action; we see the action itself. Hinshelwood then argues that the separation thesis calls this apparent epistemological datum into doubt. It is undeniable that the motions of people’s bodies are directly visible. If, as the separation thesis claims, the movements one’s body makes when one acts are not identical with one’s actions, but are instead the results of one’s actions, ‘then we might be unsure whether we really can literally see the action itself’. If someone’s arm rising is not their arm raising, then ‘[w]hat else could one see, the seeing of which would count as one’s having seen the action?’ Hinshelwood answers that ‘there is nothing else available for one to perceive’ (2013: 629–630). As I understand it, the problem that Hinshelwood identifies is that it is difficult to understand how the following statements can all be true: (a) we can directly see actions, such as someone’s raising their arm; (b) we can directly see the bodily movements that are the results of actions, such as someone’s arm going up; (c) according to the separation thesis these two things are not one

and the same. Hinshelwood thinks that the upshot is that we end up doubting that the action is really directly visible after all.

Hinshelwood acknowledges that there are other examples where we can directly see two objects that are visibly indistinguishable but nevertheless distinct. The most famous case is that of a bronze statue and the lump of bronze from which it is made. The statue and the lump are visibly indistinguishable; nevertheless, when we look in their direction we are looking at two objects, not one. The statue and the lump must be distinct objects because they each have different modal properties. The statue cannot survive being melted down, whereas the lump can. By Leibniz's law, if X and Y have different properties, then X and Y are not identical. In this case, we do not doubt the visibility of either the statue or the lump. It is not puzzling to say that there are two visibly indistinguishable, spatiotemporally coincident objects and both are directly visible because when we see one we are seeing the other.

Why, then, does Hinshelwood think it is puzzling to make a similar claim about actions and the movements one's body makes when one acts? Why can we not simply say that there are two visibly indistinguishable, spatiotemporally coincident eventualities (an action and a bodily movement₁) and both are directly visible because when we see one we are seeing the other? Hinshelwood's answer is because the two cases are not exactly analogous, and hence the action case can be puzzling even while the statue–lump case is not. In the statue–lump case we understand how the statue and lump can both be directly visible even though they are distinct by explaining that the lump *constitutes* the statue. It is understandable how we see one when we see the other because the one *constitutes* the other. If we wanted to explain how it is that someone's action and the movement their body makes when they act are both directly visible even though they are distinct, we would have to posit a relation similar to constitution to underpin their spatiotemporal coincidence and visual indistinguishability. Hinshelwood argues that it is doubtful that a relation of constitution holds between actions conceived of as causings and bodily movements₁ as the latter are supposed to be the causal results of the former.

Helen Steward (2013) offers a counterargument. She argues that, actually, it is not constitution that helps us understand how the statue and lump are both directly visible despite being distinct. What does the explanatory work here, according to Steward, are facts about how we individuate things. We understand that the statue and lump cannot be one and the same because of Leibniz's law. Because the statue and the lump have different modal properties, we understand that they cannot be identical. Actions and bodily movements₁ also have different properties, Steward suggests. Actions are things that are done; bodily movements₁ are not. Actions can be, for example, eager; bodily movements₁ cannot. This is sufficient to explain how actions and bodily movements₁ can be distinct even though they are visually indistinguishable.

However, I do not think this reply succeeds. This is because the puzzle that needs explaining is not how two things can be distinct despite being visually

indistinguishable. The puzzle is *how two things can both be directly visible* if they are distinct. What needs explaining is how when we see one we see the other. It is not enough to be reassured that the two objects are really distinct despite their visual indistinguishability. We need some explanation of what underpins their visual indistinguishability that we can use as reassurance that we really can directly see them both. It seems to me that it really is constitution, and not Leibniz's law, that explains that puzzle.

Thus, one reason to doubt the separation thesis is that it opens up a challenge to explain how it is possible that actions and bodily movements_i can both be directly visible even though they are distinct. Of course, this puzzle may be solvable—just as it is in the statue–lump case. Even if constitution is not the right way to solve it, proponents of the actions-as-causings view may be able to give some other account of the relation between actions and bodily movements_i that makes it clear how when see one we see the other.

The second reason I find the separation thesis counterintuitive is because of what it seems to imply about our relationship to our own bodies. Adrian Haddock (2005) suggests that, if the separation thesis is true, then persons are alienated from their bodily movements. According to Haddock, if the separation thesis is true, then 'our bodies are pictured as entities whose powers are wholly distinct from our powers of agency, as entities that we can (at best) only cause to move—and in this respect they are the same as any other worldly object' (2005: 161). I am not sure the separation thesis entails something quite as strong as that. The separation thesis does not, for instance, entail that moving my body is not a basic action. It does not entail that in order to move my body I must first do something else, as I have to do when I want to move other worldly objects: to move them, I need to move my body first. The separation thesis does not, therefore, collapse this distinction between moving our bodies and moving other worldly objects. Furthermore, on the view we are currently considering, actions are causings of bodily movements—they are not events that are the causes of bodily movements. This means that there is a difference between moving a glass of water and moving my arm in order to pick up a glass of water because in the first case we could say that the movement of the glass of water is caused by a prior event that I cause, i.e. the movement of my arm, whereas in the second case we cannot say that the movement of my arm is caused by a prior event that I cause, because the movement of my arm *is* what I cause. Therefore, I do not think it is correct to say that the separation thesis entails that bodies are treated 'the same as any other worldly object'.

However, I agree with the general discomfort Haddock expresses. The separation thesis says that what happens with my body when I act is not my action; it is instead the result of my action. This, to me, implies that my action constitutes my executive supervision, as it were, of what goes on with my body. What goes on with my body would not happen without me—I am the cause of my bodily movements_p, after all—but I am somewhat pulled back from what is happening with my body.

Indeed, Steward describes the control agents have over their own bodies using the metaphor of a supervisor (2012: 51, 52, 68, 162, 165). However, I do not think that we are present in our bodies as supervisors. To me, the separation thesis has parallels with Cartesian dualism. Instead of thinking of ourselves as one thing that is both physical and capable of apparently non-physical activities such as thinking, Descartes concluded from his meditations that we must be two separate substances joined together: a body and a mind. Descartes posited an additional entity—a mind—to be that which thinks, rather than accept that some physical things might be capable of non-physical activities. The separation thesis strikes me as similar in some ways. The separation thesis posits an additional entity—a causing—to be the exercise of our agential power, rather than accept that some bodily events are exercises of our agential power. Also, like Descartes's mind–body distinction, the separation thesis distinguishes our agency into a personal and bodily aspect. I am uncomfortable with this distinction between ourselves and our bodies. I agree with Haddock that the powers of our bodies are not wholly distinct from our powers of agency. We have the agential powers that we have only because of what our bodies are capable of. For example, I can lift things because of the power of my brain to stimulate my muscles and the power of my muscles to move my bones etc. My intuition is that the connection between ourselves and our bodies is much closer than that of supervisor and supervisee.

5.2.2 *A response to Alvarez and Hyman*

One way to prove the innocence of both the separation thesis and the idea that actions can be events is to show that, actually, both ideas are consistent with accepting that agency ought to be understood in terms of agent causation. I think that Alvarez and Hyman's argument for the separation thesis and for the conclusion that actions are not events is invalid. Alvarez and Hyman's argument is invalid because it wrongly assumes that the expression 'caused to rise' means 'caused an arm-rising event to happen.' Alvarez and Hyman rightly claim that it is implausible 'that causing an event to occur is not merely an event itself, but the very same event as the event caused' (Alvarez & Hyman 1998: 229). However, this only falsifies the claim that my causing my arm to rise is my arm's rising if 'causing my arm to rise' is taken to mean 'caused an arm-rising event to happen.' But why should we 'relationalise' the infinitival phrase 'causing my arm to rise'? Why should we assume that what claims like 'the agent caused her arm to rise' mean is that an agent is the cause of an arm-rising event? Rowland Stout rightly points out that '[t]he phrase "your arm to rise" is not really a noun phrase at all and certainly does not encode some implicit reference to an entity which is the event of your arm's rising' (2010: 104). In other words, the *language* we use to talk about what an agent causes when they act does not entail the metaphysical conclusion that when an agent raises her arm, a relation

of causation comes to obtain between the agent and an arm-rising event. The thesis that agency consists in an agent coming to stand in a causal relation to an event is a substantive metaphysical thesis—it is not simply what phrases like ‘the agent caused her arm to rise’ mean.

Ursula Coope (2007) outlines a response to Alvarez and Hyman’s (1998) argument that is available to Aristotle, who also thought that my arm’s going up, the arm-rising event, was identical with my action of raising my arm. Coope suggests that Aristotle would deny that his view commits him to the implausible idea that the causing of an event is one and the same as the event caused, because Aristotle would deny that an action is a causing of an event to happen. According to Coope’s Aristotle, an action is the causing of a state to obtain:

Aristotle’s view, I shall argue, is that the power that is exercised in an action of moving X is a power to produce the end of X’s movement: a power to produce a state, rather than a movement. In this sense, what I am causing when I move X is the state that X’s movement is directed towards. For example, when I raise my arm, what I am causing is my arm’s being up, rather than my arm’s going up. More generally, the action of changing something towards being F is, for Aristotle, a particular kind of causing of the state being F. (2007: 113–114)

However, another more radical response to Alvarez and Hyman is available. Suppose we rejected the relational interpretation of ‘causing my arm to rise’. Suppose we thought that an agent’s causing her arm to rise does not entail that the agent stands in a causal relation to anything. So not only does an agent raising her arm not stand in a causal relation to an arm-rising event; she also does not stand in a causal relation to the state of her arm being up. Now it is possible to accept that my causing my arm to rise is my arm’s rising, and that actions are events, because this no longer entails the absurd claim that my causing an event to happen is the event caused.

The temptation to assume that claims like ‘the agent caused her arm to rise’ mean an agent is the cause of an arm-rising event is a consequence of an incomplete rejection of relationalism. Relationalism says that causation is always and everywhere a relation between distinct entities (‘cause’ and ‘effect’). Those who endorse the actions-as-causings view reject an event-causal theory of agency and so reject the idea that what it is for a person to act can be analysed in terms of some kind of relation between two events. However, they still seek to explain agency in terms of a causal relation. Agent causation is understood in relational terms: it is taken to be a relation of causation that obtains between an agent and an event. As a consequence of this partial rejection of relationalism, actions are construed as the-coming-to-obtain of a causal relation between an agent and an event *and not* as events themselves. The positive view I will advance in the following chapters involves the complete rejection of relationalism, which

allows me to retain much of what seems right about the actions-as-causings view, without also having to accept the separation thesis or that actions are not events.

References

- Alvarez, M and Hyman, J 1998 Agents and their actions. *Philosophy*, 73(284): 219–245. DOI: <https://doi.org/10.1017/s0031819198000199>
- American Honda Motor Co. Ltd. Public Relations Division 2007 ASIMO Technical Manual, September 2007. Available at <https://asimo.honda.com/downloads/pdf/asimo-technical-information.pdf> [Last accessed 3 September 2023].
- Bach, K 1980 Actions are not events. *Mind*, 89(353): 114–120. DOI: <https://doi.org/10.1093/mind/lxxxix.353.114>
- Chisholm, R 1964 Human freedom and the self. In Kane, R *Free will*. Malden, MA: Wiley Blackwell.
- Chisholm, R 1976 *Person and object: A metaphysical study*. London: Routledge.
- Coope, U 2007 Aristotle on action. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 81: 109–138. DOI: <https://doi.org/10.1111/j.1467-8349.2007.00153.x>
- 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 2001 pp. 43–62.
- Davidson, D 2001 *Essays on actions and events*. 2nd ed. Oxford: Clarendon Press.
- Haddock, A 2005 At one with our actions, but at two with our bodies: Hornsby's account of action. *Philosophical Explorations*, 8(2): 157–172. DOI: <https://doi.org/10.1080/13869790500095939>
- Harré, R and Madden, E H 1975 *Causal powers*. Oxford: Blackwell.
- Hinshelwood, A 2013 The metaphysics and epistemology of settling: Some Anscombean reservations. *Inquiry: An Interdisciplinary Journal of Philosophy*, 56(6): 625–638. DOI: <https://doi.org/10.1080/0020174x.2013.841044>
- Hornsby, J 1980 *Actions*. London: Routledge.
- Hornsby, J 2004 Agency and actions. In: Steward, H and Hyman, J *Agency and action*. Cambridge: Cambridge University Press. pp. 1–23.
- Hyman, J 2015 *Action, knowledge, and will*. New York: Oxford University Press.
- Lowe, E J 2008 *Personal agency: The metaphysics of mind and action*. New York: Oxford University Press.
- Mayr, E 2011 *Understanding human agency*. New York: Oxford University Press.
- O'Connor, T 2000 *Persons and causes: The metaphysics of free will*. New York: Oxford University Press.

- O'Connor, T 2009 Agent-causal power. In: Handfield, T *Dispositions and Causes*. Oxford: Oxford University Press.
- Pereboom, D 2014 *Free will, agency, and meaning in life*. New York: Oxford University Press.
- Reid, T 1788 *Essays on the active powers of man*. Edinburgh: John Bell, Parliament-Square, and London: G G J & J Robinson.
- Steward, H 2012 *A metaphysics for freedom*. Oxford: Oxford University Press.
- Steward, H 2013 Responses. *Inquiry: An Interdisciplinary Journal of Philosophy*, 56(6): 681–706. DOI: <https://doi.org/10.1080/0020174x.2013.841055>
- Stout, R 2010 What are you causing in acting? In: Aguilar, J H and Buckareff, A A *Causing human actions: New perspectives on the causal theory of action*. Cambridge, MA: MIT Press.
- Taylor, R 1966 *Action and purpose*. Englewood Cliffs, NJ: Prentice-Hall.
- Turri, J 2018 Exceptionalist naturalism: Human agency and the causal order. *Quarterly Journal of Experimental Psychology (Hove)*, 71(2): 96–410. DOI: <https://doi.org/10.1080/17470218.2016.1251472>
- von Wright, G H 1962 On promises. *Theoria*, 28(3): 277–297.
- von Wright, G H 1971 *Explanation and understanding*. Ithaca, NY: Cornell University Press.