

CHAPTER 9

A New Theory of Intentional Action

One of the main aims of this book is to explain how physicalism, causal theories of intentional action and a relational approach to causation are linked. I argued in the first half of this book that these three theoretical positions are mutually supporting and form what I called the physicalist triad. I argued that we have good reason to reject the physicalist triad because the picture of human agency the triad entails is inadequate. The chief failing of the physicalist/event-causal account of agency is that it eliminates the agent from the causality of her action, which contradicts an essential part of our concept of agency—that the agent herself brings about changes. This is known as the disappearing agent objection. Agent-causal accounts of agency avoid the disappearing agent objection as they construe agency as a kind of causation where the agent exercises causal power and this exercise of causal power cannot be reduced to causation by an event involving the agent. However, I argued in Chapter 5 that agent-causal accounts face a number of issues because the metaphysical assumptions about causation they rely on are not sufficiently distinct from the relational approach to causation.

In the second half of this book, I started to navigate a path out of the physicalist triad. In Chapter 6 I proposed a non-relational theory of causation. According to this theory, causation is not always a relation but can be a process that substances engage in. Chapters 7 and 8 were concerned with explaining how this non-relational theory of causation allows us to challenge the standard causal theory of action explanation. The non-relational theory of causation allows us to think of rationalising explanations as providing causal information even though the concepts employed in such explanations do not designate causes of the actions they explain. This has consequences for how we ought to understand what intentional action is. The task of the present chapter is to make good on my promise that a non-relational theory of causation, and the

How to cite this book chapter:

White, Andrea. 2024. *Understanding Mental Causation*. Pp. 175–195. York: White Rose University Press. DOI: <https://doi.org/10.22599/White.j>. License: CC BY-NC 4.0

ontology that permits it, supports an alternative view of intentional action. I propose that to act intentionally is to engage in a process, and as such is to exercise a power—but a power of a special sort. The power to act intentionally is a power to structure one's own activities so that they demonstrate a pattern—a pattern that is only revealed by attributing mental states to the agent.

9.1 A neo-Aristotelian theory of agency

Before turning my attention to *intentional* action, it is necessary to say something about what action in general is. Part of the task of philosophy of action is to explain what agency is, or what it is to act. Like other agent-causal accounts, I propose that we understand agency in terms of substance causation. Like other agent-causationists, I believe that 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. However, my account of what substance causation is differs from standard agent-causal accounts. In Chapter 6, I outlined a distinctive non-relational understanding of substance causation that made use of a novel process ontology. I said that processes are universals and can be described as ways for substances to be changing, to be effecting change or to be resisting change. Processes that are (to some degree) ways for substances to be effecting change are species of causation. These mostly active processes I will call *activities*. What it is for a substance to be causing something is for there to be an *activity* that the substance is engaging in. A substance engaging in an activity is an agent, and the event that results once the substance has completed the activity it has been engaging in is an action. Actions are thus events of a special kind: they are events that are instances of activities.

Importantly, agents are not causally related to their actions. Individual actions are events that come into existence when an agent engages in an activity and then completes that activity. So understood, actions are 'produced by' or 'brought into being by' agents, but the sense of production here is ontological not causal. This metaphysics of action distinguishes my account from standard agent-causal accounts, which take substance causation to be a relation between a substance and an event. It also helps us see why the causality of action is something that essentially involves the agent (and thereby avoids the disappearing agent problem). On my theory, the causation exemplified by actions is the activity the agent engages in; it is something that goes on, but only insofar as it is engaged in by an agent. Furthermore, the dynamic state of affairs that is an activity going on is something that is partially constituted by the agent. A dynamic state of affairs is, as I proposed in Chapter 6, a complex entity composed of a substance and a process. So, if we take the causality of action to be a dynamic state of affairs, then the agent herself *partially constitutes* the causality

of action—she cannot, therefore, be merely the arena within which the causation of her action takes place.

Hornsby has described views like mine as ‘neo-Aristotelian’:

Neo-Aristotelians do not treat cause as everywhere a relation—neither as a relation between two events, nor between two objects, nor between an object and an event ... They take an object’s powers to tell us what kinds of processes the object can engage in, so that they connect our understanding of causality with our recognition of the display of the potentialities of things by the things having those potentialities. Thus they defend a metaphysics in which a substance ontology belongs, and to which such notions as powers, capacities, liabilities are central ... Causality, then, is present in the world inasmuch as something is actually exercising its powers, perhaps affecting something else in doing so. (2015: 131–132)

The theory I have just proposed tells us what sort of entity an action is (an event, i.e. an instance of activity). My theory also tells us what sort of entity the exercise of power is: the exercise of power by a substance is a dynamic state of affairs, i.e. a substance’s engaging in a process. However, providing a *metaphysics of action* is not all that is required for a complete and adequate theory of *agency*. It takes more to provide an adequate theory of agency than simply to describe the ontological structure of the worldly entities that are picked out by the concepts of *action*, *agent* and *activity*. To provide a complete theory of agency, one must consider the concept of agency and provide some sort of dissection of this concept.

I believe there are two distinctions crucial to our concept of agency: the distinction between activity and passivity, and the distinction between one-way and two-way powers. Agency cannot be identified with either the exercise of active power or with the exercise of two-way power. Instead, *both* concepts are key to understanding agency. The agency concept has something to do with the idea of agents as things that bring about change. John Hyman suggests that ‘to act is to intervene, to make a difference, to make something happen, to cause some kind of change’ (2015: 33). Agents *cause* change and should be contrasted with patients, who *undergo* or *suffer* change (Hyman 2015: 34). On this understanding of agency, plants, animals and inanimate objects can be agents. They are agents whenever they cause something to happen. I agree with Hyman that the concept of *agent* is kindred with *causation*, *production* and *activity*, so the notion of active power is essential to understanding what agency is. It might sound strange to say that inanimate objects can be agents but denying that inanimate objects can act is at odds with the language we use to report actions. We typically report actions by means of causative verbs like ‘melt’, ‘burn’ and ‘pump’. But we say things like ‘the acid melted the beaker’, ‘the poker burnt the cloth’ and

‘his heart pumped blood’ just as readily as we say ‘the cook melted the butter,’ ‘the criminal burnt the evidence’ and ‘the man pumped the water.’ As Hyman (2015: 30–31) has argued, it is implausible to think that these verbs have different meanings when they are used to report what inanimate things have done and when they are used to report what human beings have done.

Even though I think it literally true that inanimate objects can be agents, and they are agents when they exercise active power, there is more to the concept of agency than activity. Agency and activity are not synonyms. It seems to me that one has not really mastered the concept of agency until one has recognised the difference between things that lie there until something else comes along and prods them into action, and things that, sometimes with effort, move themselves about. It seems to be an essential part of our concept of agency that acting must involve a very minimal kind of autonomy.

Agency is connected to the idea of being able to move oneself. It contrasts with what Aristotle called ‘moved-movement.’ Therefore there is an important difference between the agency of inanimate objects and the agency of animals and human beings—and understanding this difference is essential to understanding the agency concept. This is because, as well as being kindred with concepts like *causation*, *agency* is associated with ethical concepts like *responsibility* and *blameworthiness*. As Hyman (2015) puts it, some instantiations of agency have an ‘ethical dimension’ as well as a ‘physical dimension.’ It is of great ethical significance that some actions are up to the agent whereas others are not up to the agent. There is an important moral difference between pushing someone over when you could have refrained from doing so and pushing someone over because someone else pushed you into them. This distinction has something to do with agency, and I think the terms ‘settling,’ ‘self-movement,’ ‘up-to-us-ness’ and ‘origination’ are all different ways philosophers have attempted to describe this crucial contrast. I think the best way to understand this contrast is using the concept of a two-way power.

I endorse Kim Frost’s definition of a two-way power as one that has ‘two fundamental, mutually exclusive kinds of exercise,’ whereas a one-way power has only one fundamental kind of exercise (2013: 612). The easiest way to spell out this idea is by means of an example. In the right circumstances my power to sing is two-way. What this means is that, if I do end up singing, I am manifesting my two-way power, but if I end up *not* singing (which might involve actively doing something else, but might not—it might involve continuing an activity already in progress, or letting something happen to me), I am *also* manifesting my two-way power. Thus, my power to sing, because it is two-way, is sometimes manifested by singing, and sometimes manifested by *not* singing. The power has two mutually exclusive kinds of exercise, which I will call positive and negative, and only one of these (the positive) is the activity the power is specified as a power to do.

In the case of one-way powers, when the conditions are right for the manifestation of a one-way power, the activity the power is a power to do will be

engaged in, whereas in the case of two-way powers, when the conditions are right for the *positive* manifestation of a two-way power, the two-way power may *not* be exercised positively—it may be exercised negatively—and thus the activity the power is a power to do may not be engaged in. It is important to note that, while one-way powers can be distinguished into those that are active and those that are passive, the active–passive distinction does not have application in the case of two-way powers. This is because two-way powers are powers to act *or refrain*, so they are all powers to be active in a certain way, or not (which might be to be active in a different way, or might be to be passive).

Steward (2013a) finds the conception of two-way powers as powers with two distinct fundamental kinds of manifestation problematic. For Steward, a power to ϕ is two-way just in case the agent who possesses the power to ϕ *also* possesses the power not to exercise their power to ϕ (2013a: 691). Steward argues that a conception of two-way powers like mine (and Frost's) has counterintuitive consequences (2013a: 691). As Steward notes, it seems to entail that in not singing right now while I'm working on this chapter, I am exercising my power to sing, albeit negatively. I accept that it is counterintuitive to think that, in not singing right now, I am exercising my power to sing. It is more intuitive to think that my power to sing is dormant while I am working on this chapter: it is not being exercised at all. I thus acknowledge that not *every* case where an agent does not ϕ counts as a negative exercise of a two-way power to ϕ ; not every case of not doing something is a case of *refraining* from doing it. However, I think a conception of two-way powers as powers with two mutually exclusive kinds of exercise is compatible with the fact that not every case of not doing something is a case of *refraining* from doing it.

As long as one can say something about how to distinguish cases where a two-way power to act is exercised negatively from cases where the power to act is just not exercised at all, then one is permitted to claim that there is more to exercising a two-way power to ϕ negatively than simply not ϕ ing. I doubt that there is a completely general way to distinguish cases where an agent exercises her two-way power to ϕ negatively from cases where an agent's not ϕ ing does not count as a negative exercise of her two-way power to ϕ . This is because what it takes for some instance of not acting in a certain way to count as refraining from acting in that way might depend on the type of action in question. For example, the fact that I am consciously aware of my cup of coffee might be sufficient for my not reaching for the cup to count as a negative exercise of two-way power to reach for it. But, for my not singing right now to count as a negative exercise of two-way power to sing, I may need indexical knowledge that the circumstances I am in are circumstances in which I could (or should) be singing. In all cases of refrainment, I think some sort of awareness of what one could be doing is required, but precisely what sort of awareness is required differs depending on the type of action in question.

Maria Alvarez (2013) argues that, most of the time when human beings exercise their agential powers, the power they are exercising is a two-way power.

This falls in line with the intuition that most of the time human agency is 'self-movement' and involves at least a minimal kind of autonomy. This minimal autonomy consists in our activities being up to us, in the sense that our power to perform these activities is two-way.

One challenge facing any theory of agency that appeals to two-way powers is whether this entails that agency is incompatible with determinism, the doctrine that every event is completely causally determined by prior events and conditions together with the laws of nature. Helen Steward (2012) argues that it does, and so much the worse for determinism. In other words, because agency must be understood as a two-way power, if this entails that agency is incompatible with determinism, then determinism must be false as it is undeniable that agency exists. Possibly, if one were convinced of the truth of determinism one could then argue for the non-existence of agency, just as hard incompatibilists argue for the non-existence of free will. However, denying the existence of agency seems a very high price to pay. A more plausible strategy for those convinced of the truth of determinism is to argue that possessing two-way powers is compatible with determinism.

A number of compatibilists have argued that determinism is compatible with possessing the ability to do otherwise. Some of these compatibilist arguments could be used to show that possessing two-way powers is compatible with determinism. This is because a necessary condition for having a two-way power to ϕ at a time t is to be able both to ϕ and not ϕ at t (Alvarez 2013: 108).⁴¹ The kind of compatibilist argument that could be used to show that possessing two-way powers is compatible with determinism are those that analyse the ability to do otherwise *modally*, i.e. the agent is able to do otherwise just in case it is possible for the agent to do otherwise.⁴² Compatibilist arguments that analyse the ability to do otherwise *conditionally*, i.e. the agent is able to do otherwise just in case they would have done otherwise had they tried to (or intended to, or chosen to), would not work. This is because, on a conditional analysis of the ability to do otherwise, an agent cannot possess the ability not to ϕ whenever she is able to ϕ (the necessary condition for possessing a two-way power to ϕ). Possessing the ability not to ϕ is conditional on what the agent tries/intends/chooses: if they try to ϕ at t , then they do not possess the ability not to ϕ at t . Compatibilist arguments that analyse the ability to do otherwise *modally* could be used to defend the idea that two-way powers are compatible with determinism. These

⁴¹ Another necessary condition for having a two-way power to ϕ at a time t is to have the opportunity both to ϕ and not to ϕ at t (Alvarez 2013: 108). If agent A has the ability to ϕ , then she has the right attributes for ϕ ing and knows how to ϕ (for example, A only has the ability to wave her arms if she has arms and knows how to wave them). If A has the opportunity to ϕ , then there is nothing preventing her from ϕ ing (for example, she is not tied up or injured). See also: Kenny (1975: 33).

⁴² Berofsky (2011), Campbell (2005), Kapitan (2011) and List (2014).

arguments turn on the idea that there can be more than one meaning of 'possible'. This allows one to argue that, even if determinism entails that only a ϕ ing action at t is *physically* possible given prior events and conditions and the laws of nature, it may still be possible in another sense for the agent to not ϕ at t . For example, it could still be *agentially* possible for the agent to not ϕ at t .

I will not adjudicate on the question of whether possessing a two-way power is compatible with determinism here. The fact that it is possible to argue that agency understood as a two-way power is both compatible with determinism and incompatible with determinism suggests that perhaps agency cannot settle the question of whether determinism is true or not.

Still, recognising that human agency is often the exercise of a two-way power has several advantages.

First, it can explain why there is no intentional action in deviant causal chain cases. As mentioned in Chapter 4, deviant causal chain cases are a well-known problem for event-causal analyses of intentional action, i.e. analyses that attempt to reduce intentional action to causation of bodily movements by appropriate mental states and/or events. The most famous deviant causal chain case is Davidson's (1973/2001: 79) example of a climber whose desire to rid himself of the weight of carrying another man and belief that he could do so by loosening his hold cause him to become so nervous that he lets go unintentionally.

For the event-causal theorist there is no intentional action in this case because the causal chain does not follow the sort of causal path that counts as 'the "right" way in which beliefs and desires must yield behaviour for genuine intentional action to occur' (Bishop 1989: 135), the 'right way' being '...', where the '...' has to be filled in without reference to intentional action. The success of this explanation depends on how the '...' is filled in and, as we saw in Chapter 4, no account of how the '...' ought to be filled in has been completely counterexample-free.

An alternative explanation is made available if we assume that exercising a two-way power is necessary for intentional action. If possessing and exercising a two-way power is a necessary condition for acting intentionally, then there is no intentional action in deviant causal chain cases because the agent's reasons or intentions or mental states rob the agent of the relevant two-way power, most probably by robbing the agent of the opportunity to both ϕ and not ϕ . For example, in Davidson's example, the climber's nervousness robs the climber of the opportunity not to let go of the rope. Just as extreme grief can render a person incapable of not crying out, the climber's control over his body has been hijacked by the conditions responsible for his nervous state. It is no longer up to him whether he lets go or not.

We can also now explain why some heteromesial cases are such that intentional action is blocked, and others do not block intentional action: not every heteromesial case is such that the agent is stripped of either the ability to ϕ or not ϕ or the opportunity to ϕ or not ϕ . When Amy is using her device just to keep my neural systems in working order, she has not robbed me of the ability

or opportunity to not make tea, which is why I am still exercising agency in that example, whereas where she uses her machine to control the movements of my body she has robbed me of the opportunity not to make tea.

Another advantage of explaining agency in terms of two-way powers is that we can now explain how agency can be demonstrated in passivity as well as in activity. When one's agential power is two-way, one can demonstrate this power by *not* performing the action one's agential power is a power to do. For example, in cases of intentional refrainment, e.g. where I let my plant die by not watering it or allow a telephone to continue ringing by not answering it, the putative agent exercises a two-way power to act negatively. In failing to water my plant, I do not *actively* cause the death of the plant. Substances in the vicinity that might have actively caused the death of the plant probably include parts of the plant itself (e.g. the plant's chloroplasts may have actively caused the death of the plant by using up what water was stored in the plant, thereby causing the plant to wilt, which in turn prevented the plant from capturing light etc.). In this case, I demonstrate agency by letting the active powers of *other* substances manifest themselves, rather than by exercising any active powers myself. In this case, I possess a two-way power to water the plant and I exercise my power to water the plant *negatively*. In Hyman's example of a child allowing themselves to be picked up, the child is demonstrating agency because the child is manifesting her two-way power to resist being picked up (e.g. by pushing away the parent) negatively. So, even though the child is, so to speak, not doing anything but rather letting something happen to her, she is demonstrating an agential power.

I also think that using *both* the distinction between active and passive powers and the distinction between one-way and two-way powers to explain what agency is has a distinctive advantage. The question 'what marks the difference between things that one does, and things that befall one?' is a complicated question. It is complicated because there are lots of different distinctions that have a bearing on it: the distinction between causing change and suffering change; the distinction between automatic behaviours and intentional ones; the distinction between moving oneself and being moved to move by something else. Appealing to both active and passive powers and one-way and two-way powers can help clarify this question. Agency does not reduce to the exercise of active power, because some substances can manifest their agency by remaining passive, and therefore by not engaging in activity. Neither does agency reduce to the exercise of two-way power, because not all substances that cause things to happen do so by exercising two-way powers, but all substances that cause things to happen are agents. My view is that agency is a complex, highly abstract concept that incorporates both distinctions. Some substances' agential powers are one-way; these substances manifest their agency when they are active but not when they are passive. For these substances, exercising their agential power is to engage in an activity. Other substances' agential powers are two-way; these substances manifest their agency when they are active but also

sometimes when they are passive. For these substances, in *some* cases exercising their agential power is to engage in an activity, but in other cases exercising their agential power is to allow other substances to act upon them.

Understanding agency using both the active–passive distinction and the distinction between one-way and two-way powers also has the advantage of giving us more conceptual resources for discussing some of the tricky cases discussed in Chapter 4, including reflexes, sub-intentional action, and spontaneous expressions of emotion.

Most of us would agree that reflexes, like blinking or blushing or sneezing or the knee-jerk reflex, are not intentional. Opinions are more divided on the question of whether reflexes are genuine actions. We call them ‘reflex actions’ and they are things that we do. However, they are not activities over which we have any kind of control. It is not up to me whether or not I blink when an object touches my eye; when a doctor hits my patella tendon with a reflex hammer I cannot but move my leg. For this reason, it seems wrong to attribute reflex actions to the person. Instead, reflex actions are more properly attributable to sub-personal systems. They are controlled by neurons in the spinal column and lower parts of the brain. When we perform reflex actions, we seem to be ‘moved-movers.’ We are moved to move by sub-personal systems. When we perform reflex actions, we are like ASIMO: our movements are strictly governed by our component parts.

Using the two distinctions that I believe are crucial to understanding agency, we can explain why reflex actions are called actions and described as things that we do even though it would be wrong to think of them as genuine demonstrations of human agency. Reflex actions can sometimes count as exercises of causal power. Suppose I kicked over and broke a vase as a result of stimulation of the knee-jerk reflex. In this case, I caused the vase to break and so I exercised a causal power. I was active rather than passive with respect to the breaking of the vase (though I was passive with respect to moving my leg—I didn’t get my leg to move; the doctor and my own sub-personal systems did), so in a sense I was the agent of the vase’s breaking. However, the power I exercised here was one-way and not two-way. I could not have refrained. Human beings are the kind of creatures whose movements are often up to them, hence we are the kind of creatures whose agential powers are two-way. Given this, reflex actions are not genuine demonstrations of human agency because they are not exercises of two-way power.

What about sub-intentional actions and spontaneous expressions of emotion? In Chapter 4, I described these as examples of agency that were nevertheless not intentional. The reasons I outlined for counting these as examples of agency were (a) because they are attributable to the person and not to another agent or sub-personal system; (b) because it is natural to speak of the person moving their body in cases of sub-intentional action and spontaneous expressions of emotion—in other words, they seem to be examples of self-movement; and (c) sub-intentional actions and spontaneous expressions of emotion seem

to be behaviours over which we are in control. I can now add that these examples count as demonstrations of agency because they are exercises of two-way power. When I absent-mindedly fiddle or tap my feet to music, I have the ability and opportunity not to engage in that behaviour, and that is what my control over the activity consists in. When I spontaneously embrace a loved one or laugh at a joke, again, I have the ability and opportunity not to, which is why it is true to say that engaging in these activities is up to me.

One could object to the idea that sub-intentional actions and spontaneous expressions of emotion are exercises of two-way powers. Alvarez (2013: 113) lists spontaneous expressions of emotion as actions that ‘we cannot generally avoid doing’ and thus as counterexamples to the thesis that human agency involves the exercise of two-way power. One could perhaps say the same about sub-intentional actions—they are activities we cannot generally avoid doing. However, Alvarez offers a good response. She suggests that spontaneous expressions of emotion (and presumably sub-intentional actions too) lie on a continuum that ranges from out of our control and attributable to sub-personal systems, to under our control and attributable to us. Another way of putting this point is to say that *some* spontaneous expressions of emotion are really out of our control and attributable to sub-personal systems, whereas others are within our control and attributable to us as persons, and some fall in between these two extremes. Alvarez further suggests that these activities will seem closer to one or the other end of this continuum to the extent that we are aware of our doing them. The more aware we are, the more able we are to control the activity. Alvarez then argues that whether an activity falls towards the ‘controlled by sub-personal systems’ end of the continuum or towards the ‘controlled by us’ end of the continuum depends on ‘the extent to which we determine when they happen, suppress them if we choose ... that is, to the extent to which doing them involves exercising a two-way causal power to move’ (2013: 114).

I agree with many of Alvarez’s suggestions. I agree that spontaneous expressions of emotion, and sub-intentional actions, fall onto a continuum between attributable to sub-personal systems and demonstrations of our own agency. However, I disagree with Alvarez’s suggestion that doing something can be an exercise of two-way power to a greater or lesser extent. Whether or not an activity is the exercise of a two-way power seems to me to be a binary property, not something that can come in degrees. Nevertheless, I still think that the concept of two-way powers can be helpful in this case. Most of the things that we do necessitate performing a number of sub-activities. For example, to tap my foot I need to contract certain muscles in my leg. Depending on how strongly I contract these muscles I can vary how vigorously I tap my foot. Similarly, to laugh I might contract my diaphragm as well as muscles in my face and abdomen, and I can control the quality of my laughter by controlling these various contractions. There is variability in how many of these sub-activities are exercises of two-way power. Sometimes they all are. If I’m

paying particularly close attention, or if I am very skilled, I can control not only whether or not I tap my foot but also the exact manner in which I do so.⁴³ Sometimes, only the macro-activity is an exercise of two-way power. In this case, it could be up to me whether or not I tap my foot but not up to me exactly how I do this. (Unskilled movements are often like this.) I also think that, sometimes, the macro-activity might not be under our control but the detail might be. That is, sometimes it might not be up to me whether or not I tap my foot or laugh but it *is* up to me exactly how I do it. My suggestion is that the greater the number of sub-activities that are exercises of two-way power, the more inclined we are to say that the macro-activity is attributable to the person and not to sub-personal systems.

9.2 Intentional action

I now turn my attention to the nature of intentional action. The causal theory of action maintains that intentional actions are events. On this point, I agree. Most versions of the causal theory of action maintain that at least basic intentional actions are bodily movements. For example, the action of raising my arm is one and the same event as my arm's rising (Davidson 1987: 37). On this point, I also agree. However, this is not yet a complete answer to the question of what intentional actions are, as not all bodily movements are intentional. To complete the story, the causal theory of action maintains that events count as intentional actions when and only when they are caused, in the right way, by mental states of the agent that also rationalise the action.⁴⁴

Much of what has been presented in Chapters 6, 7 and 8 points to the conclusion that construing intentional action as events caused to happen by mental antecedents is not the right way to understand intentionality. I propose an

⁴³ I do not think that attention and awareness is always what makes the difference here. For example, professional ballet dancers can control muscles in their feet that non-dancers would not be able to control. Professional dancers are therefore able to complete a wider array of very precise movements with their feet, which are necessary for being able to dance *en pointe*, for example. I would say that a professional ballet dancer can control the exact manner of her foot movements when dancing *en pointe*—that each of these finer movements was up to her—even though, while she is dancing, it is very unlikely that she is paying attention to them; she is much more likely to be thinking about what she is trying to express through her dancing. It seems to me that many highly skilled movements are like this: many of the sub-activities are up to the agent, but the agent does not need to attend to them to execute them with control.

⁴⁴ See Bishop (1989: 40–44), Davidson (1963/2001: 3–21; 1971/2001: 43–63), Mele (2003) and Smith (2012).

alternative view of intentional action. To act intentionally is to engage in a process, and as such is to exercise a power—but a power of a special sort. The power to act intentionally is a power to structure one's own activities so that they demonstrate a pattern—a pattern that is only revealed by attributing mental states to the agent. So, when an agent acts intentionally, they engage in the process of causation. The process they engage in counts as *mental* causation in virtue of the fact that the agent is manifesting a special power to organise their activities so that they instantiate a certain structure, a structure that is made comprehensible by the agent's mental states. This account builds on an account offered by Erasmus Mayr (2011).

9.2.1 Mayr's theory of intentional action

Mayr (2011) offers a theory of intentional action that takes seriously the idea that intentional action is the manifestation of a special sort of power. According to Mayr, 'intentional behaviour displays a certain characteristic structure of "purposefulness"' (2011: 271). Mayr proposes that to act for a reason is for one's behaviour to display a particular kind of structure, i.e. 'the characteristic structure of taking something as one's "standard of success and failure", or "of correctness and incorrectness"' (2011: 271). Mayr takes this proposal to be supported by the fact that, when searching for a rationalising explanation of someone's action, the facts we consider relevant are facts about whether the agent's behaviour, feelings and reasoning display—or would display—a certain pattern. For example, when we wonder if Beth is buying flour because she wants to make bread, we seek to find out things like "will Beth also buy yeast?", "if Beth got home and found out her bread tin was missing, would she feel disappointed?" and "would Beth make use of her desire to make bread in a practical deliberation?" For Mayr these facts do not merely constitute the epistemic criteria for determining what reason an agent acted in light of, they are also the facts that *make it the case* that an agent acted for a specific reason. There's nothing more to acting for a reason than for this welter of facts concerning the agent's actual or hypothetical behaviour and thinking to obtain.

What are the facts the obtaining of which makes it the case that an agent acted for a specific reason? According to Mayr's theory, there are three sorts:

1. Facts concerning the teleological structure or 'plasticity'⁴⁵ of the agent's actual or hypothetical behaviour. Mayr claims that, when an agent has a certain goal, they will 'react sensitively to changes in the environment which threaten the attainment of that goal or make it otherwise necessary to adopt different means for attaining his goal' (2011: 271)—or

⁴⁵ Mayr takes 'plasticity' to be an alternative term, used by Woodfield (1976), for this pattern in an agent's activity.

would if such environmental changes occurred. Agents with a goal will take ‘corrective measures’ and perform actions ‘conducive to overcoming obstacles’ should such mistakes or obstacles occur (2011: 271). These ‘corrective movements’ indicate to an observer that the agent has a ‘standard by which—at least implicitly—he assesses his behaviour and considers himself—in cases of non-conformity of his behaviour to this standard—to have “made a mistake” (2011: 273). When an agent does not encounter any obstacles or make any mistakes, the agent’s actions may not display plasticity. Mayr insists that, in this case, ‘our ascriptions of aims rely on our confidence that certain counterfactual conditionals about what the agent would do if obstacles arose are true, and that the hypothetical behaviour he would display would have an adequate teleological structure’ (2011: 274). In other words, the plasticity of hypothetical as well as actual behaviour is important.

2. Facts concerning the agent’s actual and hypothetical success and failure feelings. Achieving one’s aim is often accompanied by feelings of satisfaction or joy, and failing to achieve one’s aim is associated with feelings of disappointment or frustration. For Mayr, what occurrences trigger (or would trigger) feelings of satisfaction or disappointment are important for determining what the agent is aiming at, or what the agent considers to be a success and what he considers to be a failure. Of course, success is not always accompanied by feelings of joy, and failure is not always accompanied by feelings of frustration. For example, when one achieves something one considers a necessary evil, one may feel bitter and unhappy upon achieving it. In such cases, Mayr thinks that ‘the only success feeling of the agent may be a half-hearted or even bitter feeling of “having done it” or “being finished”’ (2011: 277).
3. Facts concerning whether the agent makes use of their purported aim as a premise in the practical deliberation leading to the action, or at least would if practical deliberation were called for. According to Mayr, when an agent is guided by the requirements he takes to be placed on him by his aims, this guidance will express itself in ‘individual or joint practical deliberation about what to do, before or during the action, and in *ex post* justifications of his actions. In practical deliberation, the purpose provides the premise in the agent’s deliberation, from which he proceeds to the conclusion that he should act in this way; and after the action it is to this aim that he appeals in justifying his action (as far as he is sincere)’ (2011: 279).

According to Mayr, an agent’s behaviour displays the structure characteristic of ‘purposefulness’ when facts of these three sorts obtain. Mayr claims that it is not necessary that facts of *all* three sorts obtain for an agent to act for a reason. Mayr thinks that sometimes an agent may not deliberate about what to do before acting, may be at a loss when asked later why he acted as he did,

have no success and failure feelings, and yet still act for a reason. For example, someone who has an unconscious (or subconscious) desire to sabotage a rival might give them bad advice. In this case, the agent has an aim (to sabotage his rival), but does not deliberate, would not be able to give an *ex post* justification for his action, and might not feel satisfied once the sabotage has been achieved. According to Mayr, ‘what is present in such cases is only the (actual or hypothetical) teleological structure of the agent’s behaviour’ (2011: 282). Mayr thinks this indicates that facts of the first type are privileged in the sense that where an agent is acting with an aim, facts of the first type *must* obtain—something that doesn’t hold true for the second or third type of facts.

9.2.2 Expanding on Mayr’s theory

There are two issues with Mayr’s account I would like to discuss. First, not all intentional activities display a pattern as sophisticated as the one Mayr describes. Some intentional actions are not done for reasons. For example, when I skip just for the fun of it, I have no aim I want to achieve by skipping. In such cases, because I have no aim I want to achieve, I have no aim to use in practical deliberation. Furthermore, because there’s nothing I want to achieve by skipping, there are no success or failure feelings.⁴⁶ It is also unclear that I would engage in actions that are conducive to overcoming obstacles when I skip just for the fun of it. When I skip just for fun, it is more than likely that should some obstacle to skipping occur—e.g. my path becomes blocked or dangerously slippery—I would just stop skipping. I am doing it just for fun after all, not to achieve anything, so I have no motivation to continue skipping when doing so becomes difficult. Similarly, some animal behaviour seems to be intentional, in a minimal sense, even though it does not display anything as sophisticated as Mayr’s ‘plasticity’. For example, it seems to me that, when a cat grooms itself, the grooming is intentional, but it doesn’t seem that, had the cat’s environment presented an obstacle to grooming—e.g. had it started to rain—the cat would try to overcome this obstacle and continue grooming itself. In such circumstances, the cat is as likely to run off and hunt for mice as it is to go inside and continue grooming itself there. Many animal actions are, I think, intentional, but few have as sophisticated a teleological structure as Mayr describes.

Second, Mayr endorses the idea that rationalising explanations ‘explain actions by making them intelligible’ and not by positing an event-causal link between the agent’s action and an appropriate mental event (2011: 269). What’s

⁴⁶ If I go to skip and suddenly find myself unable, this will no doubt incur negative feelings, but they are not obviously ‘failure feelings’—I am more likely to feel surprised and possibly concerned that a skill I thought I had has suddenly disappeared!

more, Mayr seems to endorse a context-placing or structural view of rationalising explanations:

When we understand acting for a reason as following a standard of success ... it must be the function of reasons-explanations to locate the action within the structure constituted by the agent's behaviour, emotional responses, thoughts, and practical reasoning which is constitutive for following the relevant standard of success. (2011: 292)

Mayr thus agrees with Julia Tanney that rationalising explanations explain by situating an agent's action within a wider pattern of activity the agent is engaging in, which thereby makes the action expected. Mayr also seems to agree with Megan Fritts that rationalising explanations are a form of structural explanation: rationalising explanations explain by connecting the agent's action with a more far-reaching description of the agent's activities in the same way in which structural explanations connect an explanandum with the form of the system of which the explanandum is part.

However, Mayr also thinks that rationalising explanations are a kind of disposition-citing explanation (2011: 295). He claims that, when a rationalising explanation is offered, a 'certain item of behaviour is explained *as the manifestation of one of the dispositions* connected with the welter of material and counterfactual conditionals which are responsible for the characteristic structure of intentional agency' (2011: 294, emphasis added). Mayr claims that the power manifested in intentional action is a 'complex power to act in certain ways in specific situations'; it is a power *of the agent* to structure her own activities (which are exercises of her abilities to act), a power that is 'superimposed on the pre-existing active powers of the agent' (2011: 295). So, on Mayr's view, rationalising explanations do two things: (a) they place the action explained within a specific structure and (b) they explain an action as the manifestation of a special sort of power to structure one's own activities, a power that is 'superimposed' on the pre-existing active powers of the agent. The second issue with Mayr's account I want to draw attention to concerns how rationalising explanations can perform both roles, and where this special power of an agent to structure her own activities comes from.

In response to the first issue, one might simply insist that actions like skipping for the fun of it and animal actions are not intentional because they do not meet the criteria Mayr sets out. However, even though actions like skipping for the fun of it and animal actions do not display a teleological structure as complex as the one Mayr describes, it is not true that they display no teleological structure at all. Anyone who can skip is able to make all sorts of small adjustments to their movements to maintain balance, or to ensure that the steps and hops that constitute skipping are executed with the required coordination. Skipping still involves some 'corrective measures', albeit on a smaller scale than the kind of corrective measures Mayr talks about. Similarly,

when a cat grooms itself, it must coordinate the movements of its body so that its tongue catches its fur in just the right way. Again, there is a form of teleological structure demonstrated. In both cases, there is a pattern demonstrated by the agent's actions—a pattern that makes sense once one learns what the agent is trying to do. I think that it is more in keeping with Mayr's core claim, that what makes an activity intentional is its characteristic structure of 'purposefulness', to grant that actions like skipping for the fun of it and animal actions are intentional in virtue of the teleological structure they display than to insist that such actions do not count as intentional because they fail to demonstrate a teleological structure of the right level of sophistication. If we are content to depart from traditional theories of intentional action and instead adopt a theory that ties the intentionality of some activity to the plasticity of that activity, then why not also accept the phenomenon of intentionality itself is not a homogenous phenomenon but instead something that can be more or less sophisticated?

The difficulty with weakening Mayr's view so that all activities that display some degree of plasticity count as intentional is that plasticity can be displayed in the behaviour of things that do not really act intentionally, for example machines and robots. This difficulty parallels issues surrounding Daniel Dennett's (1987) intentional stance theory. Dennett proposed that treating objects as rational agents with beliefs and desires helps us understand and predict the behaviour of those objects. Treating objects as rational agents with beliefs and desires is to take an intentional stance with respect to that object. According to Dennett, 'any object—or as I shall say, any *system*—whose behaviour is well predicted by this strategy is in the fullest sense of the word a believer' (1987: 15). Dennett goes on: '*What it is to be a true believer is to be an intentional system, a system whose behaviour is reliably and voluminously predictable via the intentional strategy*' (1987: 15, emphasis in original). The problem with Dennett's theory is that we can take the intentional stance to objects that do not really have beliefs and desires, like machines and robots.

It is commonly thought that there is a difference between *really* believing something and behaving *as if* you believed something, and that the difference lies in there being something extra, something hidden, in the case of genuine belief. I think this is the wrong way to capture the difference. True, machines and robots do not really have beliefs and desires, but this is not because believing something is a peculiar kind of property, or involves engaging in a peculiar kind of process. Rather, it is because machines and robots do not possess and exercise *two-way powers*. Their behaviour is not up to them. There is a real difference between behaviour of machines that seem to instantiate a pattern that can be made sense of by attributing mental states and genuine intentional action, but the difference does not consist in there being something *extra* present in the latter case. The difference is that machines are not capable of intentional action, because they do not possess two-way powers,

and possessing and exercising a two-way power is a necessary condition for acting intentionally.

A consideration that supports the idea that intentional agency always involves the exercise of two-way power is the fact that when an agent is constrained so that they only have the opportunity to ϕ , and lack the opportunity to not ϕ , if the agent ϕ s in this situation we wouldn't want to say they ϕ ed intentionally.⁴⁷ For example, suppose Ben's hands have been temporarily paralysed so that he is denied the opportunity to move his hands. Whether Ben moves his hands or not is not up to him. Is it possible for Ben, in this situation, to intentionally refrain from moving his hands? Suppose someone unaware of Ben's situation said to him, "If you keep your hands perfectly still I'll give you £10." Ben may want to comply but, even if not moving is what Ben wants, it does not seem like he is remaining still intentionally when his hands are paralysed. It seems like being able to both move and not move your hands is a precondition for doing one or the other intentionally, and lacking this two-way power renders intentionally doing one or the other action impossible.

Another consideration that speaks in favour of the view that exercising a two-way power is a necessary condition for intentional action are cases of deviant causation. As discussed above, if possessing and exercising a two-way power is a necessary condition for acting intentionally, then we can explain why there is no intentional action in deviant causal chain cases; in such cases an agent's mental states rob the agent of the relevant two-way power.

The idea that possessing and exercising a two-way power is a necessary condition for acting intentionally suggests a possible answer to the second problem facing Mayr's account. It is because we have two-way powers that our activities can demonstrate patterns of the kind Mayr describes. When we have two-way powers, it is up to us whether we perform the activities these two-way powers are powers to do. In virtue of this, the pattern our actions display is also up to us. This is where, I think, the special power of an agent to structure her own activities, the power that Mayr says is 'superimposed' on the pre-existing active powers of the agent, comes from. Because we have *many* two-way powers, we also have an extra power to organise our actions in such a way so as to meet our aims. The power to act intentionally is thus an emergent power—a

⁴⁷ Frankfurt cases (Frankfurt 1969) are thought to demonstrate that this claim is false, that an agent can intentionally ϕ , and indeed be morally responsible for ϕ ing, even when they could not have done otherwise. However, I would argue that even in Frankfurt cases the agents in question do, in fact, have the ability and opportunity not to ϕ . The presence of neuroscientists with fancy machinery, who could take control over an agent's body just in case they start to look like they might not ϕ by themselves, may foreclose the physical possibility that a ϕ ing won't happen, but these facts are not relevant to what is an open agential possibility for the agent.

power that emerges from our possessing two-way powers to act. Having such a power does not mean we will always use it—many exercises of two-way powers are not intentional, for example absent-minded fiddling. The power may also come in degrees: creatures whose powers are mostly two-way will be able to organise their activities into a greater variety of patterns than creatures whose powers are mostly one-way. This allows us to articulate one way in which human action and animal action differ. Human beings possess more two-way powers than animals, which is to say that a greater proportion of human agential powers are two-way. This allows human beings to organise their activities into more complex patterns to meet a wider variety of aims.

This view has interesting consequences for the question of what causal information rationalising explanations provide. First, the view grants that rationalising explanations are a form of disposition-citing explanation. Intentional actions are manifestations of a special sort of power, namely a power to organise one's activities in accordance with a certain form (a power that depends on having two-way powers to act), and the function of rationalising explanations is to tell us which form the agent was disposed to structure her activities in accordance with. In this way, rationalising explanations tell us that the agent's activities are manifestations of a disposition to engage in activities that fall within a certain structure. For example, "Beth is buying flour because she wants to make bread" tells us that Beth's flour-buying is a manifestation of a disposition to engage in activities that are conducive to making bread, i.e. of her special power to organise her activities in accordance with a pattern that will be deemed successful by Beth if it ends with a loaf of bread.

Second, the view allows that rationalising explanations are also context-placing or structural. On the view proposed, intentional actions are manifestations of a special power of agents to organise their activities into a pattern of determinate form. As Mayr proposes, to act intentionally is for one's behaviour to display a particular kind of structure, i.e. 'the characteristic structure of taking something as one's "standard of success and failure", or "of correctness and incorrectness"' (2011: 271). The mental concepts cited in rationalising explanations make the structure of an agent's intentional activity intelligible. When we explain Beth's buying flour by attributing to her a desire to make bread, the function of this mental concept is to show how Beth's buying flour is part of a larger pattern of activities that display the structure typically associated with 'wanting to make bread'. When we learn that Beth is buying flour because she wants to make bread, we learn that Beth's activity sits within a pattern of activity that might include buying yeast, feeling disappointed if the bread tin is missing, consulting a cookbook etc.

Third, the view can explain why determining whether rationalising explanations provide information relevant to the manipulation or control of an effect, and hence whether rationalising explanations are causal, is difficult. As I mentioned in Chapter 8, when you learn that some agent's activity is a

manifestation of her desire or an output of her rational capabilities, you learn that you might be able to alter her activity by altering what she believes about the world, or by changing her desires, perhaps by changing her environment, but more usually by reasoning with her, talking to her, or persuading her. However, learning this information only makes it the case that you *might* be able to alter the agent's activity. The view of intentional action sketched in this section allows us to explain why this is: reasoning with an agent in an attempt to prevent them from ϕ ing (or to get them to ϕ) doesn't take away the agent's two-way power to ϕ . Because her power to ϕ is two-way, it is up to her whether she ϕ s or not. Of course, we can always control someone else's ϕ ing by removing their two-way power to ϕ , for example by tying them down so that they no longer have the opportunity to ϕ . But learning about the reasons and motives behind an agent's activity is not relevant for our exercising *this* kind of control over the agent. If learning about the reasons and motives behind an agent's activity is relevant for the manipulation or control of their behaviour at all, then it is relevant for a kind of control that leaves the agent's two-way powers intact.

Determining whether rationalising explanations provide information relevant to the manipulation or control of an effect is difficult because it is unclear whether this latter sort of control is a form of *causal* control. Is convincing someone to behave in some way to exercise a *causal* power? Is it to cause something to happen? These questions matter if, as I have proposed, an explanation is causal if and only if it provides information relevant to manipulation and control, where manipulation and control are causal activities that powerful particulars, such as ourselves, can undertake. I do not think that the causal concept sits comfortably with concepts like *convincing*, *persuading* and *reasoning with*. On the other hand, the concept does not feel wholly inappropriate either. In short, because the disposition manifested when an agent acts intentionally is one which is dependent on their having and exercising *two-way* powers, learning about the reasons and motives behind an agent's activity does not provide us with information that enables us to *ensure* that the activity is (or is not) engaged in. However, it is not obvious that exercising causal control over a situation is always a matter of *ensuring* certain outcomes. The causal status of rationalising explanations is atypical. But if something like the account of intentional action I have sketched in this section is true, then the unique causal nature of rationalising explanations is not an anomaly; it is instead something that should be expected given the nature of the agential powers demonstrated in intentional action.

In this chapter, I have proposed an alternative view of intentional actions, inspired by Mayr (2011), which takes intentional actions to be manifestations of a special power of agents to organise their activities into a pattern of determinate form (an emergent power that depends on one possessing two-way powers to act). Rationalising explanations reveal this form by attributing mental states with certain contents to the agent.

References⁴⁸

- 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>
- Berofsky, B 2011 Compatibilism without Frankfurt: Dispositional analyses of free will. In: Kane, R *The Oxford handbook of free will*. 2nd ed. Oxford: Oxford University Press. pp. 153–174.
- Bishop, J 1989 *Natural agency: An essay on the causal theory of action*. Cambridge: Cambridge University Press.
- Campbell, J K 2005 Compatibilist alternatives. *Canadian Journal of Philosophy*, 35(3): 387–406. DOI: <https://doi.org/10.1080/00455091.2005.10716595>
- Davidson, D 1963 Actions, reasons, and causes. *Journal of Philosophy*, 60(23): 685–700. DOI: <https://doi.org/10.2307/2023177>. Reprinted in Davidson 2001 pp. 3–20.
- 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 1973 Freedom to act. In: Honderich, T *Essays on freedom of action*. New York: Routledge and Kegan Paul. pp. 137–156. Reprinted in Davidson 2001 pp. 63–82.
- Davidson, D 1987 Problems in the explanation of action. In Smart, J J C et al. *Metaphysics and morality: Essays in honour of J.J.C. Smart*. New York: Blackwell. pp. 35–49.
- Davidson, D 2001 *Essays on actions and events*. 2nd ed. Oxford: Clarendon Press.
- Dennett, D 1987 *The intentional stance*. Cambridge, MA: MIT Press.
- Frankfurt, H 1969 Alternate possibilities and moral responsibility. *Journal of Philosophy*, 66(23): 829–839. DOI: <https://doi.org/10.2307/2023833>
- Frost, K 2013 Action as the exercise of a two-way power. *Inquiry: An Interdisciplinary Journal of Philosophy*, 56(6): 611–624. DOI: <https://doi.org/10.1080/0020174x.2013.841043>
- Hornsby, J 2015 Causality and ‘the mental’. *HUMANA.MENTE Journal of Philosophical Studies*, 8(29): 125–140.
- Hyman, J 2015 *Action, knowledge, and will*. New York: Oxford University Press.
- Kapitan, T 2011 A compatibilist reply to the consequence argument. In: Kane, R *The Oxford Handbook of Free Will*. 2nd ed. Oxford: Oxford University Press. pp. 131–150.
- Kenny, A 1975 *Will, freedom and power*. Oxford: Basil Blackwell.

⁴⁸ Author note: some references to Davidson are formatted (1963/2001). This indicates the initial date of publication of the paper (in this case 1963) but references the paper as it appears in the 2001 collection of his essays, with the page numbers relating to that volume.

- List, C 2014 Free will, determinism, and the possibility of doing otherwise, *Noûs*, 48(1): 156–178. DOI: <https://doi.org/10.1111/nous.12019>
- Mayr, E 2011 *Understanding human agency*. New York: Oxford University Press.
- Mele, A 2003 *Motivation and agency*. Oxford: Oxford University Press.
- 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 2012 *A metaphysics for freedom*. Oxford: Oxford University Press.
- Steward, H 2013a Responses. *Inquiry: An Interdisciplinary Journal of Philosophy*, 56(6): 681–706. DOI: <https://doi.org/10.1080/0020174x.2013.841055>
- Woodfield, A 1976 *Teleology*. Cambridge: Cambridge University Press.