

## CHAPTER 2

# Causal Theories of Intentional Action

In this chapter I turn my attention to the second element of the physicalist triad: causal theories of intentional action. Two central questions within philosophy of action are ‘how do reasons explain actions?’ and ‘what is the nature of intentional action?’. The two questions are related, as part of what makes intentional actions distinctive is that often (but not always) when we explain an intentional action, that is, say why the agent acted as she did, we do so by giving the agent’s reason for acting as she did.<sup>8</sup> Explanations that cite an agent’s reasons are called ‘rationalising explanations.’ Rationalising explanations explain why an agent acted as she did (this is the explanandum) by telling us why, in the agent’s eyes, what she did was a rational thing for her to do (this is the explanans). The nature of intentional action is thus inseparable from intentional action’s appropriateness for receiving rationalising explanations. Whatever intentional actions are, they must be things that can be explained by reasons.

The first question concerns *how* rationalising explanations explain. How does a statement telling us why what an agent did seemed to them to be rational explain why the agent did as she did? How does the explanans of a rationalising explanation illuminate the explanandum? An influential answer to this question is the answer offered by Donald Davidson. Davidson (1963) argued that rationalising explanations are causal explanations. Davidson claimed that the explanantia of rationalising explanations are facts about what the agent wants to do (or what the agent has an urge to do, or what the agent has an ambition to do) and facts about what the agent believes about how to do it. Davidson calls

---

<sup>8</sup> Two examples of an explanation of an intentional action that do *not* cite the agent’s reasons or motives are: ‘Sally bit the policeman because she was drunk’ (Hyman 2015: 105) and ‘She threw the water at him because she was angry at him.’

---

### How to cite this book chapter:

White, Andrea. 2024. *Understanding Mental Causation*. Pp. 39–54. York: White Rose University Press. DOI: <https://doi.org/10.22599/White.c>. License: CC BY-NC 4.0

the composite of a desire to perform some type of action and a belief about how performance of that action may be achieved ‘the *primary reason* why the agent performed the action’ (1963/2001: 4, emphasis in original). Davidson argued that, when we say the agent acted as she did *because* she wanted to do something, or *because* she believed something was the case, this ‘because’ implies causality. From this, Davidson concluded that states of desiring and states of believing—or, at least, events suitably related to states of desiring and states of believing, such as the onset of the desire or the onset of the belief—are causes of the actions they explain. Davidson’s view is commonly called the *causal theory of action explanation*.

The second question concerns what intentional actions are. One answer is that they are events, and basic actions (i.e. actions not done *by* doing something else) 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). This is not yet a complete answer, as not all bodily movements are intentional actions. Epileptic fits are bodily movements but they are not intentional. To complete the story, several philosophers have suggested that bodily movements 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 (e.g. Bishop 1989; Davidson 1963; Davidson 1971; Mele 2003; Smith 2012). This answer has become the standard account of intentional action and is commonly called *the causal theory of action*.

In the previous chapter I argued that the relational understanding of mental causation, which plays a pivotal role in arguments for physicalism, is made to seem indispensable because of implicit acceptance of these causal theories of intentional action. Many physicalists believe that intentional or voluntary human action is only possible if mental items stand in causal relations to physical events such as bodily movements. In Sections 2.1 and 2.2 I will explain in more detail how causal theories of intentional action reinforce physicalism about mentality, which will help us see that the best strategy for resisting physicalism about mentality will involve challenging key aspects of the causal theories of intentional action. In Section 2.3 I will explain how causal theories of intentional action are themselves supported by relational assumptions about the nature of causation.

## 2.1 Rationalising explanations, mental concepts and mental causation

On the causal theory of action explanation, rationalising explanations explain by giving a causal account of the agent’s action. That is, a statement telling us why what an agent did seemed to them to be a rational thing to do explains why the agent did as she did by giving us causal information. Davidson’s (1963) argument for this position is best thought of as a challenge to anyone who thinks

that rationalising explanations are not causal, as Davidson does not offer any positive reason to think that they are.

In brief, Davidson's argument is as follows. *Some* statements that tell us why what an agent did seemed to them to be rational do *not* explain why the agent did as she did. This kind of statement could be called a 'mere rationalisation'. Mere rationalisations are similar to rationalising explanations in that they also tell us why the course of action taken by the agent seemed, to the agent, to be a rational course of action to take. However, mere rationalisations do not tell us *why an agent acted as she did*—they *only* tell us why what the agent did seemed, to the agent, to be a rational thing for them to do. For example, imagine that Anna is deciding whether or not to speak at a conference. She knows that speaking at a conference will be good for her career, but in the end, she decides to speak at the conference because it will draw praise from her friends, and not because it will be good for her career (perhaps she does not really care about her career). Anna actually spoke at the conference because she would get praise from her friends, not because it would be good for her career. In this context, the following statement would be a mere rationalisation of Anna's action:

- (a) Speaking at the conference seemed rational to Anna because it would be good for her career.

This is a mere rationalisation because it explains why speaking at the conference seemed to Anna to be a rational thing for her to do—but it does not explain why Anna actually spoke at the conference. It is not true that Anna spoke at the conference *because* she thought it would help her career. On the other hand, it is true that Anna spoke at the conference because she would receive praise from her friends. That Anna would receive praise from her friends if she spoke at the conference *does* explain why Anna acted as she did. Because some statements which tell us why what an agent did seemed to them to be rational do not explain why the agent did as she did, those statements that do both must achieve this by doing more than simply revealing why what an agent did seemed to them to be a rational thing to do. And if the extra thing rationalising explanations do is not revealing causal information, what is it? This question has come to be known as 'Davidson's challenge' and Davidson thinks there is no satisfactory answer to it.

Jonathan Dancy (2000) denies that successful rationalising explanations do more than reveal why what an agent did seemed to them to be a rational thing to do. The difference between statements that rationalise but do not explain and statements that rationalise *and* explain is simply that, in the former, the belief/desire mentioned is not the belief/desire the agent acted in the light of, and in the latter the belief/desire mentioned *is* the belief/desire the agent acted in the light of. Davidson insists that the explanatory connection between beliefs/desires an agent acts in light of and the agent's action cannot be primitive—it has to hold in virtue of some other connection between the agent's beliefs/

desires and their action. But, Dancy objects, Davidson provides no argument against the following view:

[T]he difference between those reasons for which the agent did in fact act and those for which he might have acted but did not is not a difference in causal role at all. It is just the difference between the considerations in the light of which he acted and other considerations he took to favour acting as he did but which were not in fact ones in the light of which he decided to do what he did. (2000: 163)

In other words, Dancy doesn't think that Davidson provides any argument against taking 'acted in the light of' as primitive.

On Dancy's view, 'acted in the light of' performs the function in the case of rationalising explanations that truth plays in the case of other sorts of explanation. Like truth, 'acted in the light of' is a status capable of belonging to statements given as explanans, which is a necessary condition for their explanatoriness. For example, compare 'George is the firstborn of William and Kate' with 'George is the firstborn of Elizabeth and Philip' as putative explanans of the following explanandum: why is George heir to the throne? Both statements posit the kind of relationship that would guarantee George's being the heir to the throne, but only the first statement can genuinely explain why George is heir to the throne because only the first statement is true. There is nothing perplexing about the fact that truth can make the difference between two statements that both posit something that would make sense of the explanans. That only true statements can explain is plausibly a brute fact.

However, I think there *is* something perplexing about the fact that 'acted in the light of' also seems to be able to perform this function. That 'acted in the light of' can perform this function seems like something that needs accounting for—it does not seem like a brute fact. There must be something about statements that tell us the reason the agent 'acted in the light of' that grounds their explanatoriness. The question Davidson's challenge raises is: *why* does learning that Anna's reason for acting was that she would receive praise *explain* why Anna spoke at the conference? Why does 'acted in the light of' bestow explanatory power? Julia Tanney (2009) expresses the puzzle well:

Davidson claims that it would be a mistake to conclude from the fact that placing the action in a larger pattern explains it, we now understand the sort of explanation involved, and that 'cause and effect form the sort of pattern that explain the effect in the sense of "explain" that we understand as well as any' [(1963/2001: 10)]. Davidson challenges the opponents of the causal view to identify what other pattern of explanation illustrates the relation between reason and action if they wish to sustain the claim that the pattern is not one of cause and effect. (2009: 96)

The task is to spell out what ‘pattern of explanation’ is demonstrated by rationalising explanations.

I have said that Davidson thought that the pattern of explanation demonstrated by rationalising explanations is a causal one. That is, that rationalising explanations explain by giving a *causal* account of the agent’s action. However, what is the nature of the causal information rationalising explanations are supposed to provide? This question is particularly important as it has a bearing on how we ought to understand mental causation.

Davidson’s answer is that ‘the primary reason for an action is its cause’ (1963/2001: 4). It is worth taking some time to explain what Davidson means by this. In Davidson’s view, the explanantia of rationalising explanations are facts about what the agent wants to do and facts about what the agent believes about how to do it. Davidson calls the dual possession of a desire to perform some type of action and a belief about how performance of that action may be achieved ‘the *primary reason* why the agent performed the action’ (1963/2001: 4, emphasis in original). Davidson argued that ‘For us to understand how a reason of any kind rationalises an action it is necessary and sufficient that we see, at least in essential outline, how to construct a primary reason’ (1963/2001: 4).

I think that Davidson is essentially correct on this first point. I assume that explanation is a relation between facts and only facts can explain other facts.<sup>9</sup> Furthermore, I agree that the explanatory power of rationalising explanations rests on our ability to identify facts about an agent’s desires and beliefs from the statement that rationalises the agent’s action. Of course, rationalising explanations do not typically take the form ‘agent A  $\phi$ ed because A wanted to  $\phi$  and believed that  $\psi$ ing was a way to  $\phi$ ’. Sometimes this is because it suffices to explain why someone acted as they did to only mention what the agent wanted to do. For example, in (b) Beth’s action is explained in terms of her desire only:

(b) Beth is buying flour because she wants to make bread.

We do not need to be told that Beth believes or knows that buying flour is an essential preparatory action for making bread. We take it for granted that Beth possesses this knowledge.

Other times it is sufficient only to mention what the agent believes, or knows, about how to achieve what they want to do. For example, in (c), Carlin’s action is explained in terms of his belief only:

---

<sup>9</sup> Van Fraassen (1980: 134–153) proposes a theory of explanations as answers to why-questions where both the answer and the topic of the why-question are true propositions. Raley (2007) has also defended the view that all explanation is factive. See also: Bokulich (2011), Hempel and Oppenheim (1948), Kitcher (1989) and Woodward (2003).

- (c) Carlin is adding rosemary to the sauce because he believes it will make it taste better.

We do not need to be told that Carlin wants to make the sauce taste better—we take it for granted that he wants this. Davidson’s point is not that all rationalising explanations *explicitly* give the primary reason why the agent acts but rather that, for the explanans of a rationalising explanation to illuminate the explanandum, ‘it is necessary and sufficient that we see, at least in essential outline, how to construct a primary reason’ (1963/2001: 4). That is, the explanatory power of rationalising explanations rests on our ability to construct a primary reason from any rationalising explanation.

Although I think Davidson is broadly correct in thinking that the explanantia of rationalising explanations are facts about what the agent wants and believes, there is a complication. When an agent acts for a reason, the reason for which they act is not usually a fact about the agent’s own mental states. For example:

- (d) Daniel took the A road because the motorway was shut.

In (d) Daniel’s reason is ‘that the motorway was shut’, not ‘that Daniel believed or knew that the motorway was shut’. At least, that is how things seem. How does this square with Davidson’s claim that the primary reason why an agent acts is a belief–desire pair? The best way to tackle this complication is, I think, to acknowledge that the word ‘reason’ can be used in more than one way.

First, the term can be used to denote an agent’s *reason for acting*. I follow Maria Alvarez (2010) in thinking that an agent’s reason for acting is that which makes the action a sensible or rational or good thing to do. As Alvarez puts it, an agent’s reason for acting is ‘the desirability characterisation’ the action has for the agent. As such, reasons for acting are not usually facts about an agent’s mental states. Strictly speaking, Daniel’s reason for taking the A road is not that he *wants* to get somewhere and *believes* that, because the motorway is shut, taking the A road is the only means of getting there. The good Daniel sees in taking the A road is that, given that the motorway is shut, taking the A road is the only way he can get to where he wants to go.

As well as being used to denote the desirability characterisation an action has for an agent, the word ‘reason’ can also be used as a synonym for ‘explanans’. When we give *the reason why* such and such is the case, we are providing an explanans. Reasons why are explanantia of explanations. I think Davidson’s claim that primary reasons given by rationalising explanations are belief–desire pairs is plausible only if ‘primary reason’ is taken to mean ‘primary reason why’ or ‘primary explanans’, because reasons for acting are not usually facts about the agent’s own mental states. However, I believe that primary reasons why, i.e. the primary explanantia, of rationalising explanations *are* facts about what the agent wants and believes. That is, I believe that the explanatory power of rationalising explanations rests on our ability to construct a belief–desire pair from any rationalising explanation.

We are now in a better position to clearly state what Davidson means by the claim ‘the primary reason for an action is its cause’ (1963/2001: 4). Davidson’s view is not only that rationalising explanations give causal information but that rationalising explanations are true if and only if the belief or desire which explains the action (or some mental event suitably related to the belief or desire) *stands in a causal relation to the action explained*. Davidson is making *two* claims here. First, rationalising explanations give causal information. Second, rationalising explanations are true if and only if the belief or desire which explains the action stands in a causal relation to the action explained. If Davidson is correct, then the possibility of true rationalising explanations of action entails that there must be causal relations between mental items and actions.

Construing rationalising explanations as explanations which posit an entity that is causally related to the action explained encourages us to think that concepts like *belief* and *desire* refer to mental *items*. This view, I believe, legitimises a metaphysics of mind wherein our status as minded creatures depends on the existence of mental events and states whose nature we have yet to discover and whose existence must, one way or another, be reconciled with the idea that the world is physical in all its fundamental aspects. In this way, the causal theory of action explanation creates the problem physicalism is supposed to solve. The causal theory of action explanation encourages us to accept an ontology that includes mental items whose intrinsic nature is up for discovery, which stand in causal relations to human actions. If we also assume that human actions fall under the jurisdiction of scientific causal explanation, then, unless the intrinsic nature of those mental items is, somehow, exhaustively determined by the underlying physical causes of our actions, it is hard to see how rationalising explanations can be true. To put it another way, if the causal theory of action explanation is correct, then the possibility of true rationalising explanations of action entails that there must be causal relations between mental items and actions. If we also assume that actions are physical events, then the causal theory of action explanation justifies the relational understanding of mental causation, which says that mental items stand in causal relations to physical events. And, as we saw in the previous chapter, the relational understanding of mental causation is the driving force in arguments for physicalism.

## 2.2 The causal theory of action and physicalism

The causal theory of action concerns the ontological question ‘what is the nature of intentional action?’ Although I have introduced the causal theory of action as if it were one unified theory, in fact matters are more complicated than this. There are many different theories that attempt to give a causal account of intentional action. What these many theories have in common is the commitment that acting intentionally consists in events being caused to happen by non-actional mental antecedents. However, there is plenty of room for disagreement after this commitment is accepted.

Most causal theorists believe that actions, or at least basic actions (i.e. actions we perform without having to do anything else first) are bodily movements. However, some causal theorists believe that actions are composite events such as the event of an-intention-causing-a-bodily-movement, or an event that involves neural states and bodily movements. For example, Michael Smith argues that ‘we should suppose that actions are events that begin in the brain, continue on in the nervous system and muscles, and end with the relevant events of the body’s moving’ (2021: 7).

There is also disagreement on exactly what kind of mental antecedents must cause an event to happen if it is to count as an intentional action. Following Davidson’s suggestion that ‘the primary reason for an action is its cause’ (1963/2001: 4), some causal theorists have suggested that beliefs and desires must feature in the aetiology of an event, if that event is to count as an intentional action. For example, elsewhere Smith has proposed that:

[A]ctions are those bodily movements that are caused and rationalised by a pair of mental states: a desire for some end, where ends can be thought of as ways the world could be, and a belief that something the agent can just do, namely move her body in the way to be explained, has some suitable chance of making the world the relevant way. Bodily movements that occur otherwise aren’t actions, they are mere happenings. (2004: 165)

Some causal theorists take the mental antecedent necessary for intentional action to be an intention. On this kind of view, the agent’s beliefs and desires cause the acquisition of an intention to act, which in turn triggers the behaviour that constitutes the agent’s action. John Searle (1983) argues that, for an event to count as an action, the event must be caused by a specific kind of intention, namely one that continues exerting causal influence over an agent’s behaviour even after the behaviour has begun, thereby sustaining and guiding the behaviour to ensure that it satisfies the agent’s prior motive. Berent Enç (2003) also argues that for an event E to be an action it must be caused (in the way it is normally caused) by an intention, the content of which explicitly refers to bringing about an E-type event (2003: 78–79). For Enç, deliberation about what to do is a ‘computational process ... the causal consequence of which is the formation of an intention’ that in turn causes a ‘behavioural output’ (2003: 2). Others suggest that second-order desires, like the desire to act on a particular motive (Frankfurt 1978) or the desire to act in accordance with reasons (Velleman 1992), must be part of the causal history of an event if that event is to count as an intentional action. In all these versions of the causal theory of action, mental items are assigned a causal role in bringing about an event.

Perhaps the most significant source of disagreement concerns what constitutes *the right way* for a mental item to cause a bodily movement for there to

be intentional action. Not just any causal chain from mental event to physical event is sufficient for there to be an intentional action. A necessary condition for acting intentionally is that the agent is in control of what is going on with them. It is difficult to explain what is meant by control in this context without begging the question against certain theories of action. I will have more to say on what kind of control is necessary for intentional action in later chapters, but for now it suffices to illustrate with an example what kind of control is required for intentional action.

Imagine my friend Amy really wants me to make tea, so she makes sure I am thirsty by giving me something salty to eat, puts a cup and some teabags nicely in view, then says, "Why don't you have some tea?" The conditions are right for me to make tea, but whether or not I do is still up to me. I am in control of my making tea (or not) in this case. Now, suppose Amy installs some clever machinery to manipulate my brain and nervous system and uses that to make me make tea (in the manner of the character Black from Harry Frankfurt's (1969) thought experiment). In this case, I am not in control of my movements. Amy has taken control over what goes on with me.

The causal theorist would say that the difference between these two cases, what explains why I have control in the one case but not in the other, has something to do with the causal history of my movements in each case. In the second case, where Amy manipulates my brain, the causal chain leading up to my bodily movement is not the kind of causal chain required for there to be agential control. For one thing, the causal chain does not involve my own mental states. However, it is not sufficient merely to include mental states in the causal chain leading to bodily movement. These mental states have to operate in the causal chain in the right way. For there to be intentional action, the causal chain from mental item to bodily movement must be such that it constitutes the agent's control over their action. The causal chain cannot deviate from the kind of causal chain that occurs in a normal, uncontroversial case of intentional action. Davidson gives an example of a deviant kind of causal chain:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally. (Davidson 1973/2001: 79)

In this example, the climber has an end he wants to achieve and a belief about how to achieve this end. This belief–desire pair causes a bodily movement of a type that is rationalised by the belief–desire pair, just as causal theorists allege it would in an ordinary case of intentional action. But, in this case, the climber did not let go intentionally. There is great disagreement on what kind of causal

chain from mental state to bodily movement is required for the agent to retain control over their action. I will discuss this problem, known as the problem of deviant causal chains, further in Chapter 4.

Finally, there is disagreement on what exactly the causal theory of action should be a theory of. Some versions of the causal theory of action are presented as accounts of agency in general. These are presented as theories that explain the difference between things that you do and things that befall you, or ‘between bodily movements that you are making happen and those which happen without your making them occur’ (Brent 2017: 656), or ‘between actions and things that we do when we are merely passive recipients of courses of events’ (Enç 2003: 2). Other versions of the causal theory of action take for granted the distinction between events to which a person is subject and events of which the person is the agent, and offer specifically a theory of intentional action (Mele 1992; Mele 2003) or rational agency (Bratman 2001; Velleman 1992).

These disagreements are related to important questions about the nature of agency and intentional action. Some of these questions will arise again in later chapters, either when I critically evaluate causal theories of intentional action or when I present my own account of intentional action. For now, though, most of these disagreements can be set aside. To see the connection between causal theories of action and physicalism it is the core ontological commitment of all causal theories of intentional action that we need to focus on.

All versions of the causal theory of action hold that acting intentionally consists in the right kind of event being caused to happen, in the right way, by the right kind of mental antecedents. This commitment entails that acting intentionally is nothing over and above some special kind of event causation, and that the possibility of intentional action requires that certain mental items stand in causal relations. Just like the causal theory of action explanation, the causal theory of action encourages us to accept an ontology that includes mental items which stand in causal relations.

According to most versions of the causal theory of action, what mental items cause is either a physical event such as a bodily movement or an event that is composed or realised by a bodily movement. This means that, if the causal theory of action is correct, then the existence of causal relations between mental items and physical events (or events realised by physical events) is entailed by the existence of intentional action. If the causal theory of action is correct, then Kim’s claim that ‘the possibility of human agency ... requires that our mental state have causal effects in the physical world’ is also correct. In this way, the causal theory of action serves as justification for the relational understanding of mental causation.

It is difficult to endorse the causal theory of action without also being a physicalist, as the ontological component of the causal theory of action seems to set up the conditions for the causal argument for physicalism. Alfred Mele acknowledges the connection between taking a causal perspective on intentional action and physicalism. He states that the causal perspective ‘is usually

embraced as part of a naturalistic stand on agency according to which mental items that play a causal/explanatory role in intentional conduct bear some important relation to physical states and events' (2003: 6). John Bishop also acknowledges this point:

Surely we may understand how agency is naturally possible only if we first understand how mentality may be part of nature? That this is so is entirely clear if a Causal Theory of Action is to provide the solution to the problem of natural agency because this theory holds that action consists in behaviour caused by relevant mental states. And there is problem posterior to the problem of natural agency—namely, the problem of explaining how those extra properties beyond agency as such that are required for personal moral responsibility can themselves be realised within a natural scientific ontology. (1989: 8)

In Bishop's view, a complete naturalisation of our perspective of ourselves as agents capable of rational, intentional action would require a solution to 'scepticism about understanding how minds can be part of nature' (1989: 8). However, I think Bishop mischaracterises the connection between these two projects. He presents the problem of providing a naturalistic account of the mind as 'posterior' to the problem of finding a naturalistic account of agency. This implies that the former problem is in some ways independent from the latter problem. In my view, the connection between the project of naturalising the mind and the project of naturalising agency is much closer. It is the causal theory of action that encourages us to accept an ontology of causally efficacious mental items, an ontology that then needs to be reconciled with the 'naturalistic' view of what causation in the actual world is like. In other words, the causal theory of action justifies the relational understanding of mental causation, which as we have seen is the crucial premise in arguments for physicalism. For this reason, even though it is logically possible to accept the causal theory of action without being a physicalist, in practice belief in the causal theory of action supports physicalism. Because of this connection, the strongest challenge to causal arguments for physicalism will require a critical examination of the causal theory of intentional action.

I also think it is difficult to be a physicalist without endorsing the causal theory of action. This is because both theories are thought to be consistent with naturalism, a philosophical position that eschews the existence of anything that would be regarded as an unnatural addition to the world as described by science. Physicalism assumes nothing more than a world of physical things and this ontology is thought to fit comfortably with a scientific view of the world. Bishop argues that one should endorse the causal theory of action because it promises to 'make intelligible the possibility of agency within the natural order' (1989: 10). Thus, a key motivation for adopting causal theories of intentional action is that they seem to provide a naturalistic account of intentional action.

As Enç puts it, the causal theorist's starting point is that 'by assuming nothing more than a world of material things, we can understand the nature of decisions, of intentions, of voluntary action, and the difference between actions and things that we do when we are merely passive recipients of courses of events' (2003: 2). In this way, belief in physicalism lends credence to causal theories of intentional action, because both are apparently part of a naturalistic worldview. However, to see exactly *why* causal theories of intentional action are thought to be naturalistic it is necessary to examine the connections between causal theories of intentional action and the other element of the physicalist triad: relational approaches to causation.

### 2.3 Naturalistic agency and the relational approach to causation

The causal theory of action is reductive: it says that intentional action is nothing over and above event causation. The agent's role in bringing about what she intends is reduced to causation by her mental states or events. Agential control over what goes on exists, but it is exhaustively determined by some special kind of event causation. One key draw of causal theories of intentional action is that we can achieve an adequate understanding of intentional action without countenancing the existence of irreducible agent causation.

This is good, causal theorists argue, because the idea that there is the flux of causally related events and then *there are also* agents—three-dimensional substances, persons—who interfere with this flux to bring about the events they want to see happen is antithetical to the naturalistic view of the causal world. As we saw in the previous chapter, a naturalistic view of the causal world is one that endorses the relational approach to causation. According to naturalism, causation, as it exists in reality, cannot be the exercise of power because that kind of 'necessary connexion' is ineffable and empirically unrespectable. Instead, causation must be a certain kind of relation between events. The causal theory of action thus presupposes a metaphysics where causation is always, everywhere a relation between events. This approach to causation compels the causalist to seek to understand intentional action in terms of a distinction between different types of event causation. Causal reality is nothing more than a chain of causally related events, so, if intentional agency is a causal phenomenon at all, it must be located within this worldview. If you endorse the relational approach to causation, then the causation demonstrated in intentional action must be a relation, because all causation is, and will count as mental causation if and only if at least one of the terms of that relation is a mental entity.

The relational approach to causation is also presupposed by Davidson in his discussion of whether rationalising explanations of actions are causal explanations. Recall that Davidson argues that when we say the agent acted as she did

*because* she wanted to do something, or *because* she believed that something was the case, this ‘because’ implies causality. He also concludes from this that states of desiring and states of believing—or, at least, events suitably related to states of desiring and states of believing—are causes of the actions they explain. Davidson is assuming here that, if rationalising explanations reveal causal information, the causal information they reveal is that there are mental items, which the mental concepts employed in rationalising explanations pick out, that stand in causal relations to actions.

Contemporary non-causalists, who deny that rationalising explanations are causal explanations, also make this assumption. Julia Tanney is explicit about this:

[T]he position I wish to bring back into focus says that what it is for an action to be in execution of an intention or for it to be explicable by reasons is not a matter of there being a causal relation [understood as ‘a relation between two logically and temporally distinguishable events’] between intention or reasons and action. If causation is to be thus understood the pattern in virtue of which a person’s intentions, motives or reasons explain her action is not *eo ipso* causal. (Tanney 2009: 95)

However, is it right to assume that a rationalising explanation is causal only if it posits a causal relation between an item somehow picked out by the mental concept employed in the explanation and the action explained? This assumption will seem obvious if you take a relational approach to causation. If all causation is relational, then explanations that reveal causal information will reveal information about causal relations, because what other kind of causal information is there?

Although both causalists and non-causalists assume a relational approach to causation, I think this assumption is more supportive of the Davidsonian/causalist position. This is because, although I agree with non-causalists that mental concepts like *belief* and *desire* do not seem to designate causally efficacious *items*, I think the intuition that rationalising explanations are causal is hard to resist. This means there is a strong motivation to accommodate valid points made by the non-causalists, without giving up the idea that rationalising explanations are causal.

Davidson’s anomalous monism lets one do this. Davidson thinks that mental concepts are anomalous, which is to say that they are unsuitable for inclusion in causal laws of the form: ‘there is an event-kind F, of which the cause event is a token, and an event-kind G, of which the effect event is a token, such that F events always cause G events.’ This means that Davidson can acknowledge that there are significant differences between the explanatory scheme of explanations of actions that employ mental concepts and typical, scientific causal explanations. (The latter, Davidson thinks, do imply causal laws.) For instance,

Davidson can agree with non-causalists that mental concepts do not *seem* to perform their explanatory function by designating causes, because as *mental* concepts we should not expect them to. The anomalousness of mental concepts means that the causal nature of mental states and events is not revealed when these entities are picked out by mental concepts. This does not mean, however, that the facts that make that rationalising explanation genuinely explanatory are not causal facts. As Erasmus Mayr puts it:

For Davidson, the epistemological criteria that we use for determining for which reason an agent has acted are the considerations of rationality and overall coherence among his mental states that are generally relevant for the interpretative enterprise of ‘making sense of the agent’. What makes the reasons-explanation true, however, is something completely different: the obtaining of an event-causal link between reason and action, which for Davidson must be based on a strict causal law. (2011: 269–270)

The causalist can thus argue that, even though mental concepts do not *seem* to perform their explanatory function by designating causes, rationalising explanations would not be true if mental concepts did not somehow pick out events that stand in causal relations to actions. Of course, anomalous monism might not be correct, but I think that the opposition between Davidson and non-causalists on the matter of rationalising explanations is at something of an impasse, because anomalous monism is an available position.

The causal theory of action explanation and the causal theory of action are part of what is called ‘the standard story’ of human action. The theories are intuitively plausible enough to have become the standard account of what intentional action is and how it is explained, the account other theories must be weighed against. This is so despite the fact that causal theories of intentional action suffer some significant shortcomings, which I will discuss in Chapter 4. Why do causal theories of intentional action enjoy such intuitive plausibility? I contend that causal theories of intentional action seem superior to alternatives in part because philosophers of action assume a relational approach to causation. It is very difficult to imagine an alternative understanding of the causality of intentional action if you take a relational approach to causation. If causation is always, everywhere a relation between events, it would seem that the causation demonstrated in intentional action must be a relation, because all causation is, and will count as mental causation if and only if at least one of the terms of that relation is a mental entity. Furthermore, if causation is always, everywhere a relation between events, then explanations that reveal causal information will reveal information about causal relations. Thus, if rationalising explanations are causal, then they must point to causal relations between items somehow picked out by the mental concepts employed in the explanation and the action explained.

References<sup>10</sup>

- Alvarez, M 2010 *Kinds of reasons: An essay in the philosophy of action*. Oxford: Oxford University Press.
- Bishop, J 1989 *Natural agency: An essay on the causal theory of action*. Cambridge: Cambridge University Press.
- Bokulich, A 2011 How scientific models can explain. *Synthese*, 180(1): 33–45. DOI: <https://doi.org/10.1007/s11229-009-9565-1>
- Bratman, M 2001 Two problems about human agency. *Proceedings of the Aristotelian Society*, 101(3): 309–326. DOI: <https://doi.org/10.1111/j.0066-7372.2003.00033.x>
- Brent, M 2017 Agent causation as a solution to the problem of action. *Canadian Journal of Philosophy*, 47(5): 656–673. DOI: <https://doi.org/10.1080/00455091.2017.1285643>
- Dancy, J 2000 *Practical reality*. New York: Oxford University Press.
- 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.
- Enç, B 2003 *How we act: Causes, reasons, and intentions*. New York: Oxford University Press.
- Frankfurt, H 1969 Alternate possibilities and moral responsibility. *Journal of Philosophy*, 66(23): 829–839. DOI: <https://doi.org/10.2307/2023833>
- Frankfurt, H 1978 The problem of action. *American Philosophical Quarterly*, 15(2): 157–162.
- Hempel, C and Oppenheim, P 1948 Studies in the logic of explanation. *Philosophy of Science*, 15(2): 135–175. DOI: <https://doi.org/10.1086/286983>
- Hyman, J 2015 *Action, knowledge, and will*. New York: Oxford University Press.

---

<sup>10</sup> 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.

- Kitcher, P 1989 Explanatory unification and the causal structure of the world. In Kitcher, P and Salmon, W *Scientific explanation*. Minneapolis, MN: University of Minnesota Press. pp. 410–505.
- Mayr, E 2011 *Understanding human agency*. New York: Oxford University Press.
- Mele, A 1992 *Springs of action: Understanding intentional behavior*. New York: Oxford University Press.
- Mele, A 2003 *Motivation and agency*. Oxford: Oxford University Press.
- Raley, Y 2007 The facticity of explanation and its consequences. *International Studies in the Philosophy of Science*, 21(2): 123–135. DOI: <https://doi.org/10.1080/02698590701498035>
- Searle, J 1983 *Intentionality: An essay in the philosophy of mind*. New York: Cambridge University Press.
- Smith, M 2004 The structure of orthonomy. *Royal Institute of Philosophy Supplement*, 55: 165–193. DOI: <https://doi.org/10.1017/s1358246100008675>
- 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>
- Smith, M 2021 Are actions bodily movements? *Philosophical Explorations*, 24(3): 394–407. DOI: <https://doi.org/10.1080/13869795.2021.1957205>
- Tanney, J 2009 Reasons as non-causal, context-placing explanations. In: Sandis, C *New essays on the explanation of action*. Basingstoke: Palgrave Macmillan. pp. 94–111.
- Van Fraassen, B C 1980 *The scientific image*. Oxford: Clarendon Press.
- Velleman, D J 1992 What happens when someone acts? *Mind*, 101(403): 461–481. DOI: <https://doi.org/10.7591/9781501721564-008>
- Woodward, J 2003 *Making things happen: A theory of causal explanation*. New York: Oxford University Press.