

## CHAPTER 8

# Action Explanation

In the previous chapter, I argued that it is not necessary for an explanation to be causal that its explanandum designate an effect and its explanans designate an item that is the cause of that effect. My non-relational theory of causation implies that facts about causal relations between particulars are not the only causal facts that could ground the truth of causal explanations. I suggested that some causal explanations are made true by the non-relational aspect of causal reality, that is, by facts about substances engaging in processes. In this chapter, I turn my attention back to rationalising explanations of action. Rationalising explanations of action explain why an agent acted as she did (this is the explanandum) by telling us why, in the agent's eyes, what they did was a rational thing for them to do (this is the explanans). I will argue that rationalising explanations are also causal explanations that are not made true by a pair of causally related events.

The debate concerning how we ought to understand rationalising explanations is central within philosophy of action because part of what makes intentional actions distinctive is that often when we explain an intentional action we do so by giving the agent's reason for acting. The nature of intentional action is thus inseparable from their appropriateness for receiving rationalising explanations. Whatever intentional actions are, they must be things that can be explained by reasons.

For a long time, opinion on rationalising explanations has been divided in two. There are those who endorse the causal theory of action explanation, which says that rationalising explanations explain by giving a *causal* account of the agent's action, and moreover that 'the primary reason for an action is its cause' (Davidson 1963/2001a: 4). Then there are non-causalists, who believe that the concepts cited in rationalising explanations, like *belief*, *desire* and *intention*, do not refer to items that can stand in causal relations to actions, so, when

---

### How to cite this book chapter:

White, Andrea. 2024. *Understanding Mental Causation*. Pp. 155–174. York: White Rose University Press. DOI: <https://doi.org/10.22599/White.i>. License: CC BY-NC 4.0

such concepts are employed to explain why an agent acted, they do not explain by giving a *causal* account of the agent's action.

The causal theory of action explanation is widely supported in large part due to what has become known as 'Davidson's challenge,' which runs 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 is a 'mere rationalisation.' Mere rationalisations are similar to rationalising explanations in that they 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. If Anna ends up speaking at the conference to impress her friends and not because it will be good for her career, even though she considers being good for her career to be a sound reason to speak at the conference, then (a) is a mere rationalisation of her 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. That she wanted to impress her friends, on the other hand, *does* explain why Anna acted as she did. (a') is a genuine rationalising explanation of Anna's action.

- (a') Anna spoke at the conference because she wanted to impress her friends.

Because some rationalisations do not explain why the agent did as she did, those rationalisations that do—like (a')—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 that rationalising explanations do is not revealing causal information, then what is it?

As mentioned in Chapter 2, Jonathan Dancy argues that the difference between rationalising explanations and mere rationalisations is just that the former tell us 'the considerations in the light of which he acted' and the latter only tell us '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). 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: it makes the difference between a statement that explains and a statement that does not explain. Dancy thinks the fact that 'acted in the light of' can perform this function is a brute fact. However, while I think it is plausibly a brute fact that only true statements can explain, there is something perplexing about the fact that 'acted in the light of' can perform the same sort of function truth can. Why does 'acted in the light of' bestow explanatory power? What pattern of

explanation is demonstrated when we give the reason an agent acted in the light of? For Davidson, the answer is the sort of explanatory pattern we see when we give the cause of an effect. According to Davidson's proposal, of all the reasons Anna had for speaking at the conference, that it would impress her friends is the reason that explains her action because this reason somehow identifies the cause of her action. As emphasised in Chapter 2, Davidson's view is not just that rationalising explanations explain by giving causal information—they also identify *the cause* of the action explained.

It is important to note that, even though Davidson claimed that 'the primary reason for an action is its cause' (1963/2001a: 4), strictly speaking Davidson did not think beliefs or desires were causes. Believing that something is the case and desiring to do something are not events but states. In (a') the explanans, i.e. 'she wanted to impress her friends', would be classified by Mourelatos as a state predication, not an event predication. On the assumption that causal explanations are typically explanations that tell us what event stands in a causal relation to the event whose occurrence we want to explain, the fact that the explanans of most rationalising explanations is a state predication seems to speak against classifying these explanations as causal. This difference between rationalising explanations and typical causal explanations, however, does not carry much force. Davidson acknowledged that beliefs and desires are not events but states—in fact, he thought that they were 'dispositions to behave in certain ways' (1997/2001b: 72)—and as such could not literally be causes. However, Davidson argued that *the onset* of a belief and *the onset* of a desire are events, and explanations of actions that cite beliefs and desires are explanatory if the belief or desire attribution is 'closely associated' (1963/2001a: 12) with an inner mental event such as the onset of the belief or desire that is the cause of the action explained. As Davidson puts it, 'In many cases it is not difficult at all to find events very closely associated with the primary reason. States and dispositions are not events, but the onslaught of a state or disposition is' (1963/2001a: 12). The causal relation that makes the rationalising explanation explanatory does not need to be *explicitly* reported by the rationalising explanation. As explained in the previous chapter, the Davidsonian position on stative causal explanations is that, even though states cannot literally be causes, states can be 'part of the circumstances in which the cause occurred; and mentioning that state can help to explain why the cause had the effect it did' (Child 1994: 106). Davidson allowed the relation between a causal explanation and the causally related events that make the explanation true to be opaque (Child 1994: 106), which means that reference to the events whose causal connectedness grounds the truth of a rationalising explanation need not be transparent.

Another nuance of Davidson's theory of rationalising explanations concerns how he deals with the following issue. When we causally attribute one event to another, this is usually taken to imply the existence of a law that states that 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.

However, when we say that an agent acted as she did because of the beliefs and desires she had, there is no implication that other agents with the same beliefs and desires will (or are likely to) do the same thing, or that the same agent will act in the same way when she has the same beliefs and desires on another occasion (Hart & Honoré 1985: 55). Davidson's anomalous monism allows him to concede this point without giving up the idea that rationalising explanations must be causal explanations. Davidson proposes that when a mental event and an action are causally related, these two events fall under event-kinds that feature in a causal law. This follows from Davidson's principle of the nomological character of causality: all causal relations are covered by strict deterministic laws. However, the event-kinds that feature in the causal law, which the mental event and action fall under, are *physical* kinds, not mental kinds. Furthermore, the law that covers the causal relation can *only* be stated in a language of physical kinds. Rationalising explanations do not imply lawlike regularities between mental states and actions because in giving a rationalising explanation we are picking out the cause of an action using mental kinds, and these mental kinds do not feature in any universal regularity, not even the universal regularity that covers the causal relation that the rationalising explanation owes its success to. As Davidson puts it:

The laws whose existence is required if reasons are causes of actions do not, we may be sure, deal in the concepts in which rationalisations must deal. If the causes of a class of events (actions) fall in a certain class (reasons) and there is a law to back each singular causal statement, it does not follow that there is any law connecting events classified as reasons with events classified as actions—the classifications may even be neurological, chemical, or physical. (1963/2001a: 17)

Consequently, if anomalous monism is true, we should expect that when we say that an agent acted as she did because of the beliefs and desires she had, there is no implication that other agents with the same beliefs and desires will (or are likely to) do the same thing, but this is because in giving a rationalising explanation we are picking out the cause of an action using *mental* kinds, and these mental kinds do not feature in any universal regularity.

There has recently been renewed interest in developing a non-causal account of rationalising explanations that meets Davidson's challenge. For example, Scott Sehon (2005) argues that rationalising explanations are *teleological explanations* that are irreducible to causal explanations; Julia Tanney (2009; 2013) argues that rationalising explanations are *context-placing explanations*; and Megan Fritts (2021) argues that action explanations are *structural explanations*.

In order to determine whether any of these non-causal accounts meet Davidson's challenge, it is useful to outline what the success criteria for meeting Davidson's challenge are. It is important to recognise that Davidson's challenge is not an epistemological challenge: it is not a question of how we know

which reason an agent acted in the light of. For example, Davidson is not challenging us to explain *how we know* that Anna spoke at the conference because she wanted to impress her friends and not because it would be good for her career. For Davidson, this epistemological question is answered by considering how the reason fits with the agent's general character, how the reason coheres with the agent's other beliefs and desires, whether the reason reveals the agent to be acting rationally etc. However, Davidson stresses that 'it is an error to think that, because placing the action in a larger pattern explains it, therefore we now understand the sort of explanation involved' (1963/2001a: 10). Davidson's challenge is a theoretical challenge. It can only be successfully answered with a satisfying theory of how rationalising explanations explain. Successfully meeting Davidson's challenge requires answering the following question: what is it about the reason we know is the reason for which the agent acted that *qualifies* the connection between reason and action as an explanatory connection?

The view I advance in this chapter falls somewhere in between the Davidsonian and the non-causalist view. Like most non-causalists, I agree that the concepts cited in rationalising explanations do not seem to discharge their explanatory function by denoting (even opaquely) causes of the actions they explain. However, I do not think non-causal accounts of rationalising explanations successfully meet Davidson's challenge—at least, the non-causal theories that have the best chance of meeting Davidson's challenge are also theories whose classification as 'non-causal' is questionable. The way forward is, I think, to adopt a position that sits somewhere in between non-causalism and causalism. First, in Section 8.1, I will examine considerations that motivate seeking a non-causal account of rationalising explanations. In Section 8.2 I will assess some non-causalist accounts of rationalising explanations. I will argue that the 'non-causalist' accounts of rationalising explanations that stand the best chance of meeting Davidson's challenge could be considered 'causal' after all if one takes a non-Davidsonian approach to what makes an explanation causal. In Section 8.3 I will make the case for thinking that rationalising explanations are causal explanations, but causal explanations that are unique in two important ways.

## 8.1 Mental concepts

Rationalising explanations display certain features that set them apart from typical causal explanations, like 'the patient developed cancer because he was exposed to radiation'. As already mentioned, many rationalising explanations explain an action by reference to a state of the agent, as opposed to an event involving the agent. For example in (a')–(c) the explanantia are state predications:

- (a') Anna spoke at the conference because she wanted to impress her friends.
- (b) Beth is buying flour because she wants to make bread.

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

Many rationalising explanations offer facts as explanans:

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

Some rationalising explanations include infinitival phrases as explanations:

- (e) Esther is breaking eggs to make an omelette.

Many rationalising explanations, which Michael Thompson (2008) calls ‘naïve action explanations’, explain one action in terms of another:

- (f) Fred is drilling a hole in the wall because he is hanging a picture.

Some rationalising explanations explicitly mention intentions or what the agent is trying to do. However, because rationalising explanations are so variable in form, not very much at all can be concluded about how rationalising explanations explain by considering the language of rationalising explanations. Furthermore, a core tenet of the Davidsonian view is that the events whose causal connectedness grounds the explanatory power of a rationalising explanation need not be *explicitly or transparently* referenced in the rationalising explanation. However, one motivation for seeking a non-causal theory of rationalising explanation is that the concepts employed in rationalising explanations, such as *belief, desire, intention, goal* and *attempt*, do not seem to discharge their explanatory role by designating a cause of the action they are invoked to explain—not even implicitly or opaquely. The point here is that rationalising explanations do not seem to have anything to do with finding the cause of an action.

Elizabeth Anscombe points out that, when ‘one says what desire an act was meant to satisfy, one does not identify a feeling, image or idea that precedes the act the desire explains’ (2000: 17). The desire that an act satisfies is not the ‘mental cause’ of the act in the same way that, to use Anscombe’s example, noticing a face appearing at the window might be the mental cause of one’s jumping. Anscombe defines a ‘mental cause’ as ‘what someone would describe if he were asked the specific question: what produced this action ... on your part: what did you see or hear or feel, or what ideas or images cropped up in your mind, and led up to it?’ (2000: 17–18). Giving a ‘mental cause’ of something, in the special sense of ‘mental cause’ that Anscombe has isolated, is thus to say what prior mental event *triggered* one’s action. Rationalising explanations do not seem to be like this: they do not seem to be explanations whose explanatoriness depends on them identifying, or at least suggesting, the event that triggered the action. When we explain actions by citing an agent’s beliefs or desires, we are usually not identifying something that occurred at a particular

time that triggered the action, or which moved the agent from a state of inaction to a state of action. When we seek a rationalising explanation of someone's action, what triggered the action, what event made the difference to its occurring, does not seem to matter.

Julia Tanney makes a similar point. Tanney argues that the concepts that are at work in rationalising explanations perform their explanatory role 'without designating *anything*; let alone causally efficacious states or events; let alone causally efficacious states or events whose nature awaits discovery' (2009: 100, emphasis added). Tanney claims that assuming that mental concepts designate 'logically independent, temporally antecedent, causally efficacious events' is to assume that mental concepts are 'theoretical terms' (2009: 100). A theoretical term is one that purports to refer to an event, property, state, fact or condition whose intrinsic nature is up for discovery but which causes a phenomenon to be explained. An example of such a theoretical term would be 'gene': genes are entities we posit on the grounds that their existence would explain some observable phenomena; 'gene' is a term that purports to refer to a hidden but causally efficacious entity. Tanney argues that treating mental concepts as theoretical terms 'mis-assigns the explanatory function of these concepts' (2009: 100).

The position commits us to postulating an event, unobservable to others and possibly even to the agent herself, that would, if known, provide the sought-after reason-explanation for the agent's action. In such cases, as Ryle insists, an epistemological puzzle arises as to how anyone could ever know whether a person acts for reasons or what, if she does, her reasons are, since the hypothesis is not even in principle testable. Not only do we not, in everyday situations, have access to these hidden events, but even if we were, say, to monitor the neural activity of someone's brain or access their stream of consciousness, we would never be able to set up the kinds of correlations that would establish a particular occurrence as an instance of a particular reason without already having a way of deciding whether someone acted for a particular reason in order to make the correlation (Tanney 2009: 100).

Tanney's point here is that, if we construe mental concepts as designating hidden inner causes of behaviour, then rationalising explanations become the kind of claim whose truth depends on the existence of events we have no access to and that just does not seem to be how rationalising explanations work.

One might argue that Anscombe's and Tanney's view that when we explain each other's intentional behaviour we do not do so by positing inner mental causes that produce the behaviour is just an intuition. Jerry Fodor is one among many who has the opposite intuition (1987). However, Devin Curry (2018) summarises empirical evidence from experimental psychology that seems to support Anscombe's and Tanney's interpretation of what we are doing (or what we are not doing) when we give rationalising explanations.

Curry cites evidence from within the field of 'attribution theory', the branch of psychology concerned with how people explain each other's behaviour.

Research by Bertrum Malle indicates that there are important differences between the types of explanation people give for accidental behaviour and the types of explanation people give for intentional behaviour. When explaining accidental behaviour, people are more likely to reference ‘inner causes’, treating these as one would a mechanical cause of a physical event (Malle 2004: 61). The more intentional a behaviour appears to be, the more likely people are to explain the behaviour in terms of reasons (Malle 1999: 28–31). Curry also cites research that shows that children draw a distinction between *mistakes*, which demand explanation in terms of beliefs, and *accidents*, which demand explanation in terms of causes in the physical environment (Hatano & Inagaki 2002; Schult & Wellman 1997). Further research indicates that the types of questions people ask of someone’s behaviour are different depending on whether the behaviour is perceived to be intentional or unintentional. Only for unintentional behaviour do people ask what *produced* the action (Malle, Knobe & Nelson 2007; Monroe & Malle 2017). From this research, Curry concludes that when it comes to explaining intentional behaviour people are concerned with placing the behaviour in a context that makes the behaviour understandable—people are concerned with identifying what produced behaviour only when the behaviour is perceived to be unintentional or accidental.

Interestingly, the psychological evidence is inconclusive with regard to whether people regard reason-citing explanations of intentional actions as causal or not. Curry writes that, even though people seem to treat reason-citing explanations of intentional behaviour as distinct from mechanistic explanations of accidental behaviour, previous work has shown that teleological explanations are often considered a kind of causal explanation (DiYanni & Kelemen 2005; Lombrozo & Carey 2006). What Curry’s discussion shows is that Anscombe’s intuition—that when we explain intentional behaviour we do not do so by identifying an ‘inner cause’ or ‘mental trigger’ that produced the action—cannot be easily dismissed. An assumed contrast between rationalising explanations and mechanistic explanations underlies much of our actual attributional behaviour. The empirical evidence Curry (2018) cites seems to suggest that we generally treat reason-citing explanations of intentional actions as distinct from causal or mechanical explanations, only using the latter kind of explanation when explaining unintentional or accidental actions.

Another observation that gives us reason to doubt that rationalising explanations explain by giving the cause of the explanandum is that sometimes when an agent has more than one reason for performing some action it is genuinely indeterminate which of the reasons was the reason she acted for. As Erasmus Mayr puts it, there is not always a fact of the matter about which reason an agent acted for. Consider cases where the agent has a bundle of strong motives to do X but it is not clear—even after thorough examination of his action, its circumstances and his general character—on which of these motives he has acted. We do not have to assume that our inability to decide this question rests



on merely practical grounds—that is, that there is a fact of the matter that we are unable to establish only because we lack further evidence—for it may well be that we would not even know what kind of further evidence would decide the question. Instead, we should accept that in such cases our inability may stem from the fact that these cases are truly indeterminate, because the criteria for judging whether the agent acted on a particular reason have ‘run out’ without unequivocally determining an answer (Mayr 2011: 261).

The idea that rationalising explanations explain by identifying the cause of the action is inconsistent with allowing for this kind of indeterminacy. On the Davidsonian view, any indeterminacy regarding what belief or desire the agent acted in the light of is epistemic—this is because an agent acts in the light of a belief or desire if and only if the onset of that belief or desire is the cause of the action, and the latter relation cannot be indeterminate. Of course, it could be that, when an agent has many reasons favouring a course of action, their action is causally overdetermined by these many reasons. However, it seems possible that an agent could have many reasons favouring a course of action, where none of these reasons is the reason the agent acted, and where the agent would not have acted if the case for acting was not overwhelming. For example, imagine Anna is again deciding whether or not to speak at a conference, and because the conference is quite far away Anna vows only to speak at the conference if the case for doing so seems overwhelming, where overwhelming for her means that there are at least  $n$  strong reasons favouring the action (where  $n$  is more than one). Then suppose Anna discovers  $n$  reasons for speaking at the conference, and so goes on to speak at the conference, but none of Anna’s reasons stands out as the reason for which Anna spoke at the conference. In this case, it does not seem like Anna acts in the light of just one of the many reasons favouring speaking at the conference, but it is also not plausible to describe this as a case of overdetermination by her  $n$  reasons, because it is not the case that Anna would have acted in the same way had any one of her  $n$  reasons been missing.

## 8.2 Rationalising explanations as non-causal explanations

Rationalising explanations do not seem to explain action by designating inner causes of behaviour. However, the power of Davidson’s challenge is that, if Davidson’s answer is the only satisfactory answer to the challenge, then, regardless of how rationalising explanations *seem* to function, their explanatoriness *must* be grounded by causal relations between events somehow identified by mental concepts and the actions explained. What seems to matter, then, is whether there is a successful non-causal answer to Davidson’s challenge, which is, as stated above, to explicate what is it about the reason we know is the reason for which the agent acted that *qualifies* the connection between reason and action as an explanatory connection. In the next two sections I will consider

two non-causal accounts of rationalising explanation that I think have the best chance of meeting Davidson's challenge.<sup>38</sup>

### 8.2.2 Context-placing explanations

Tanney suggests rationalising explanations ought to be understood as 'context-placing' explanations. The explanans of a rationalising explanation explains the action by placing it in a context that makes it intelligible. According to Tanney, rationalising explanations are explanations that work by giving us more information about what is going on. Tanney provides the following example of a 'context-placing explanation':

- (g) The teacher has written 'CAT' on the board because she is writing 'CAT-ALYST' on the board.

According to Tanney, the explanans does not illuminate 'any mysterious connection between the occurrences of two contingently related events—the writing of "c", "a", and "t", on the one hand and the writing of "catalyst", on the other' (2009: 98). Instead, the explanans in (g) 'serves to re-characterise what happened so that it—as newly described—is no longer puzzling' (Tanney 2009: 98). Tanney argues that rationalising explanations are all, essentially, of this kind.

Tanney's theory of rationalising explanations as context-placing is similar to a suggestion made by Anscombe that rationalising explanations 'interpret' the action explained:

To give a motive ... is to say something like "See the action in this light".  
To explain one's own actions by an account indicating a motive is to put them in a certain light. (2000: 21)

---

<sup>38</sup> The two accounts I will consider are not the only non-causal accounts of rationalising explanation available. Scott Sehon (2005) argues that rationalising explanations are *teleological explanations*, which are irreducible to causal explanations. On this account, rationalising explanations explain by making clear the aim the agent's behaviour was directed towards. Anna's action is directed towards impressing her friends, not improving her career prospects. That's why the former, but not the latter, explains her action. What makes it the case that Anna's action is directed towards impressing her friends and not towards improving her career prospects? Sehon argues that facts about what an agent is aiming to achieve are not reducible to causal facts; instead they are their own *sui generis* kind of teleological fact (see also: Löhrer & Sehon 2016; Sehon 2007: 163–165; Sehon 2010: 125). Unlike Sehon's account, the two accounts I will consider do not entail any substantial metaphysical commitments.

Mayr (2011: 269) also 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. For Mayr, rationalising explanations explain by providing us with a way of framing the agent’s actions—a way of seeing the agent’s actions as manifesting a certain pattern.

Tanney’s view also bears some similarity to a view advanced by Michael Thompson (2008). Thompson (2008) outlines a class of rationalising explanations that he calls ‘naïve action explanations’. These rationalising explanations explain one action in terms of another. (f) would be an example of such a rationalising explanation:

(f) Fred is drilling a hole in the wall because he is hanging a picture.

Tanney’s paradigm context-placing explanation also explains why an agent engaged in some activity in terms of something else the agent is doing. Thompson suggests that most of the time when we explain our intentional actions we do so by citing another activity we are engaging in, of which the action to be explained is a part. Thompson grants that not all rationalising explanations have this form but he argues all rationalising explanations depend for their success on being suitably related to a relevant naïve action explanation. Tanney, Mayr and Thompson all seem to have hit upon what is essentially the same idea: 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.

How does this theory of rationalising explanations meet Davidson’s challenge? Tanney says in cases where there are multiple reasons that could make sense of an agent’s action but only one that genuinely explains why the agent acted as she did, we may simply need to ‘probe further for a different or more far-reaching context-placing explanation that will succeed or give up the initial expectation that the action can be explained by reasons’ (2009: 100). In other words, the distinction between a mere rationalisation and a genuinely explanatory rationalising explanation is that the latter, but not the former, fits better with a more far-reaching account of the agent’s activities. So Anna spoke at the conference because she is trying to impress her friends, and not because it would be good for her career, because she is not trying to improve her career prospects. Her current action (speaking at the conference) is a constituent of her broader action of trying to impress her friends; it is not a constituent of trying to improve her career prospects because that is not something Anna is doing. This can be seen if we take a broad enough appraisal of Anna’s activities and plans. Mayr would add that Anna spoke at the conference because she is trying to impress her friends and not because it would be good for her career because Anna is following standards of success set by impressing her friends and not standards of success set by improving her career prospects—this can be seen once we take into account the fact that Anna would be pleased if her

friends were impressed by her conference talk, disappointed if they were unimpressed, and feel nothing if her career prospects improved.

This issue with Tanney's (and Mayr's) response to Davidson's challenge is that the considerations outlined above seem only to address the epistemological question of how we know which reason an agent acted in the light of. Davidson's warning that 'it is an error to think that, because placing the action in a larger pattern explains it, therefore we now understand the sort of explanation involved' (1963/2001a: 10) seems particularly applicable to the context-placing view. What we need to know is why situating an agent's action within a wider pattern of activity the agent is engaging in qualifies as an *explanation* of that action. What qualifies the connection between the agent's action and a more far-reaching description of the agent's activities as an explanatory connection? If the connection were a causal one, or somehow reducible to a causal connection, that would be an answer. This objection to the context-placing theory of rationalising explanations has been raised by Megan Fritts (2021), who argues that putting something in context, thereby making it intelligible, and explaining something are 'usually considered two different goals or activities'. I am convinced that rationalising explanations are context-placing and serve to make an agent's action intelligible in just the way that Tanney, Thompson and Mayr describe. A crucial function of rationalising explanations is that they redescribe an agent's action in such a way as to situate the agent's action into the agent's wider activities. However, I also agree with Fritts that this is not quite enough to meet Davidson's challenge. What's missing is an account of why context-placing qualifies as explaining.

### 8.2.3 Structural explanations

Fritts (2021) suggests that rationalising explanations are structural explanations. Fritts takes as her starting point the idea hit upon by Tanney, Thompson and Mayr 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. As Fritts puts it, 'if ... our reasons for action are, at bottom, other activities in which we are involved—then human activity has a nesting-doll structure where smaller actions are constituents of larger activities' (2021: 20).

Fritts then argues that because rationalising explanations are explanations that appeal to the 'nesting-doll structure' of intentional actions, as opposed to the triggers or causes of the action to be explained, this makes them structural explanations.

Fritts endorses Stuart Shapiro's (1997) definition of a structure. Shapiro defines a structure as 'the abstract form of a system' where a system is 'a collection of objects with certain relations'. A structure is the form of a system, which is to say it is something that describes 'the interrelationships among the objects'

and ignores ‘any features of them that do not affect how they relate to other objects in the system’ (Shapiro 1997: 73). Explanations that appeal to the form of a system are structural explanations. According to Fritts, rationalising explanations are explanations that appeal to the fact that the action to be explained exists within a system of interrelated activities that place constraints on what activities can/should be performed.

Sally Haslanger (2016) gives a nice example of a structural explanation:

Suppose I am playing ball with my dog. I stuff a treat into a hole in the ball and throw it for him. The ball goes over the lip of a hill and rolls down into a gully. Why did the treat end up in the gully? If we imagine the trajectory of the treat alone, from a space near my hand, through an arc in the air, then landing about an inch above the ground and moving at about that height down the hill until it stops, it would be a huge task to explain the particular events that determined each of its movements. A much easier explanation would be to point out that the treat was inserted into a ball that was thrown and rolled down the hill into the gully. In this latter explanation, we explain the behaviour of the treat by its being part of something larger whose behaviour we explain. (2016: 114)

The structural explanation for why the treat ended up in the gully has a distinctive advantage over the event-causal explanation, which is that only the latter tells us why the treat would still have landed in the gully even if Haslanger’s ball-thrower had thrown the ball slightly differently. Throwing the ball slightly higher or with slightly more force would not have made a difference, as what mattered for the treat ending up in the gully is that it was inside the ball. Haslanger suggests that the structural explanation ‘provides a better model for seeing how I could intervene to prevent the treat from ending up in the gully (not throw the ball in that direction, for example, or catch up with the ball and stop it from rolling)’ (2016: 115). This feature of structural explanations has parallels in the rationalising explanation case—lending support to Fritts’s proposal that rationalising explanations are structural.

Consider for example rationalising explanation (b): Beth is buying flour because she wants to make bread. We might be able to give an explanation of Beth’s buying flour that starts with the onset of her desire to make bread, involves brain activity and muscle movements, and ends with her at the cash register paying for flour. However, that explanation would not, on its own, tell us why Beth would still have ended up buying flour even if she had, say, driven rather than walked to the shop, or walked to a different shop, or spent a day doing something else before buying flour. However, explaining Beth’s buying flour by pointing out that it is part of Beth’s intentional bread-making activities does provide this information. Both the event-causal explanation and the structural explanation involve the mental concept ‘wanting to make bread’. The former uses this mental concept to pick out a mental event: the onset of a

desire to make bread. The latter uses this mental concept to denote the ‘nesting-doll’ structure of Beth’s activities.

Does this account of rationalising explanations meet Davidson’s challenge? The problem with Tanney’s context-placing account of rationalising explanations is that it does not tell us what qualifies the connection between the agent’s action and a more far-reaching description of the agent’s activities as an explanatory connection. Fritt’s development of the context-placing account seems to answer this question: the connection is the kind of connection we see in structural explanations—i.e. one that connects the explanandum with the form of the system of which the explanandum is part.

However, an issue with this account of rationalising explanations is whether structural explanations are really *non-causal* explanations. If you take the Davidsonian view of what a causal explanation is, then structural explanations are not causal explanations. They do not function by identifying the cause of the explanandum; therefore, they are not causal explanations. However, in the previous chapter, I sketched an alternative theory of what makes an explanation causal. I stated that causal explanations are those that provide information relevant to the manipulation of an effect. They are explanations that provide us with information about how to stop something from happening, or how to get something to happen again, or how to get it to happen in a different way (or at least information about how to make such outcomes more likely). Structural explanations seem to provide this kind of information. Haslanger’s example of a structural explanation tells us how to prevent the treat from ending up in the gully, even if it does not tell us the cause of the treat’s falling into the gully. If structural explanations give us information relevant to the manipulation of an effect, then they would count as causal explanations, even if they do not identify causes of the explanandum.<sup>39</sup>

Furthermore, the fact that rationalising explanations can provide information that would tell us how we might stop an agent from performing an action seems to be relevant to their qualifying as explanations and not mere rationalisations. A key difference between (a) and (a’) is that (a’) gives us information about how we could have persuaded Anna not to speak at the conference. Telling Anna that speaking at the conference would not be good for her career would not have made a difference to her action. However, telling Anna that speaking at the conference would not impress her friends might have prevented her action. This is one way to explain the difference between a reason Anna thought justified her action but which was not the reason for which she acted, and the reason she acted in light of. Nevertheless, even if structural explanations are a species of causal explanation, the theory that rationalising explanations are structural explanations shows that it is not necessary, to meet Davidson’s challenge, to construe rationalising explanations as somehow identifying *the cause* of the action they explain, even if we have to acknowledge that part of

---

<sup>39</sup> Skow (2018) makes a similar argument.

what makes rationalising explanations explanatory is that they provide causal information. The position we have arrived at lies in between the causalist and non-causalist views. Rationalising explanations are causal explanations, but they are not explanations that function by identifying the cause of the action they explain. Instead they explain as structural explanations do: by identifying the wider structure of the agent's activities, of which the explanandum action is a part.

### 8.3 Rationalising explanations as unique causal explanations

The debate between causalists and non-causalists is a difficult debate to adjudicate on because both sides have intuitive appeal. However, there is a way to accept the non-causalist's ideas about mental concepts and how rationalising explanations explain without giving up the intuition that rationalising explanations are causal. The assumption made by both Davidsonians and non-causalists is that an explanation is causal only if it depends for its truth on the obtaining of a causal relation. Davidsonians and non-causalists both assume that there's one sort of thing causation can be; therefore, what real-world facts an explanation can answer to does not vary according to the explanatory context. This is a key part of the relational approach to causation that I have sought to challenge. The relational approach to causation says that there is one kind of causal reality true causal explanations answer to. However, as I argued in Chapter 7, my non-relational approach to causation allows us to argue that there are some causal explanations that are not made true by a pair of causally related events. Because causation is a term that can refer to processes and dynamic states of affairs as well as difference-making relations between events, there is more than one kind of causal reality that causal explanations can answer to. As I argued in Chapter 7, explanations can be causal even when they do not necessarily imply the existence of causal relations between certain particulars.

It is possible, therefore, that the peculiar features of rationalising explanations—features that set them apart from more typical event-causal explanations—are not barriers to thinking of these explanations as causal. It is possible that rationalising explanations could be causal even though the mental concepts cited in the rationalising explanation do not designate causally efficacious items. Rationalising explanations could be the kind of causal explanation that answers to the non-relational aspect of causal reality.

This thesis, that rationalising explanations are causal explanations that are made true by the non-relational aspect of causal reality, is attractive for at least two reasons. First, it allows us to save the intuition that explaining someone's actions in terms of their beliefs and desires is to give causal information, while at the same time accepting that the mental concepts appealed to in rationalising explanations do not refer to items that stand to the action explained as cause to effect. In other words, the thesis that rationalising explanations are

causal, but made true by the non-relational aspect of causal reality, allows us to acknowledge what's intuitive about both the Davidsonian and the non-causalist views. We can agree with non-causalists like Anscombe and Tanney that when we explain an agent's action by giving their reasons we are not identifying the trigger of their action, or that which made the difference to their action occurring. Instead, we are explaining why a person (a substance) engaged in a particular activity. The action that comes about when the agent completes her activity may well have a difference-making cause—but that is not what we are interested in when we give a rationalising explanation. What's more, whether it has a difference-making cause is often not relevant to the truth of the rationalising explanation.

Second, there are similarities between rationalising explanations on the one hand and process-citing and disposition-citing explanations on the other, which lends support to the idea these three kinds of explanation belong in the same general category. Some rationalising explanations appear to be very similar to causal explanations that cite the continuous operation of causal processes. Causal explanations that cite the continuous operation of causal processes are roughly of the form: some effect occurred or is occurring, or obtained or obtains, because substance *S* is or was engaging in causal process *P*. Thompson's (2008) 'naïve action explanations' have this form, as they explain why an agent engaged in some activity in terms of something else the agent is doing. Other rationalising explanations have the form of stative explanations. If, as seems plausible, mental states like desiring, believing and knowing are dispositions, then this would make those stative rationalising explanations disposition-citing explanations. Hyman (2015: 103–132) and Mayr (2011: 295) also propose that rationalising explanations that cite mental states of the agent are disposition-citing explanations.

However, there are two ways in which rationalising explanations are unique. First, if mental states like desiring, believing and knowing are dispositions, they are not ordinary dispositions. Most dispositions are dispositions to engage in or undergo a certain specific activity or process. In contrast, having a desire to do something or achieve something (for example) disposes one to undertake whatever activities are deemed, by the agent, to be acceptably good means of achieving what one wants; to deliberately refrain from acting should *that* turn out to be an acceptably good means of achieving what one wants; to feel happy or pleased if one's desire gets satisfied or disappointed if it is frustrated; and to use one's desire as a premise in practical deliberation about what to do. Desires are not dispositions to do any one specific thing (or even any two specific things)—they are rather dispositions for one's activities to instantiate a certain pattern or goal-directedness, which is made sense of by the content of the desire. Similar claims can be made about other mental concepts. As Ryle suggests, it would be wrong to think, just because the verbs 'know' and 'believe' are 'ordinarily used dispositionally', that 'there must therefore exist one-pattern intellectual processes in which these cognitive dispositions are actualised'



(1949: 44). Rather, states of believing and states of knowing, if they are dispositions at all, are ‘dispositions the exercise of which are indefinitely heterogeneous’ (1949: 44). So, while there are some similarities between rationalising explanations on the one hand and process-citing and disposition-citing explanations on the other, it is important not to forget that rationalising explanations are unique: they are very variable in form and, even if we suppose that the mental states cited in rationalising explanations are dispositions, they are not, by any means, ordinary dispositions.

Second, rationalising explanations do not exactly provide us with information about how to stop something from happening, or how to get something to happen again, or how to get it to happen in a different way (or at least information about how to make such outcomes more likely). 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. This is because reasoning with an agent in an attempt to get them to  $\phi$ , for example, does not guarantee that the agent will  $\phi$ —it does not even ensure that it is more likely that the agent will  $\phi$ . This is because the agent can ignore you, or remain unconvinced, or even just act against her better judgement. In short, rationalising explanations do not seem to be the sort of explanations that provide us with information about how to stop something from happening, or how to get something to happen again, or how to get it to happen in a different way, or even how to make such outcomes more likely. They seem only to provide information about how we *might* stop something from happening, or get something to happen again, get it to happen in a different way, or make such outcomes more likely.

In this chapter, I have argued that explanations of intentional action that cite the agent’s reasons for acting are the kind of causal explanation that are not made true by causally related events. The most important consideration favouring this view is that it saves two strong intuitions: (a) that reason-giving explanations are causal, and (b) that the mental states cited in reason-giving explanations do not denote items that stand in causal relations to the actions they explain. This view has important consequences for how we ought to think about the nature of intentional action. As mentioned, it is commonly held that we can achieve an adequate account of what it is to act intentionally by examining the distinctive sort of explanation with which intentional actions are associated. If rationalising explanations are causal explanations that do not designate mental items that stand to the action explained as cause to effect but instead answer to non-relational causal reality, then the case for thinking intentional actions are distinguished from non-intentional actions by their mental causes is significantly weakened. However, without getting clearer on exactly what facts about dynamic states of affairs rationalising explanations could plausibly be said to answer to, it

is difficult to offer a positive account of what the distinguishing mark of intentional action is. In the next chapter, I will present a view on intentional action that grants that some rationalising explanations are disposition-citing explanations and others are structural explanations but which also respects the two ways in which rationalising explanations are unique.

## References<sup>40</sup>

- Anscombe, G E M 2000 *Intention*. Cambridge, MA: Harvard University Press. Originally published in England in 1957 by Basil Blackwell.
- Child, W 1994 *Causality, interpretation, and the mind*. New York: Oxford University Press.
- Curry, D 2018 Beliefs as inner causes: The (lack of) evidence. *Philosophical Psychology*, 31(6): 850–877. DOI: <https://doi.org/10.1080/09515089.2018.1452197>
- 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 2001a pp. 3–20.
- Davidson, D 1997 Indeterminism and antirealism, In: Kulp, C B *Realism/antirealism and epistemology*. Lanham, MD: Rowman & Littlefield. pp. 109–122. Reprinted in Davidson 2001b pp. 69–84.
- Davidson, D 2001a *Essays on actions and events*. 2nd Edition. Oxford: Clarendon Press.
- Davidson, D 2001b *Subjective, intersubjective, objective*. Oxford: Clarendon Press.
- DiYanni, C and Kelemen, D 2005 Using a bad tool with good intention: How preschoolers weigh physical and intentional cues when learning about artifacts. *Cognition*, 97: 327–335. DOI: <https://doi.org/10.1016/j.jecp.2008.05.002>
- Fodor, J 1987 *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
- Fritts, M 2021 Reasons explanations (of actions) as structural explanations. *Synthese*, 199(5–6): 12683–12704. DOI: <https://doi.org/10.1007/s11229-021-03349-4>

---

<sup>40</sup> Author note: some references to Davidson are formatted (1963/2001a). This indicates the initial date of publication of the paper (in this case 1963) but references the paper as it appears in the 2001a collection of his essays, with the page numbers relating to that volume. Similarly, some references to Davidson are formatted (1997/2001b) which indicates the initial date of publication (in this case 1997) but references the paper as it appears in the 2001b collection of Davidson's essays, with the page numbers relating to that volume.

- Hart, H L and Honoré, A M 1985 *Causation in the law*. 2nd ed. Oxford: Oxford University Press.
- Haslanger, S 2016 What is a (social) structural explanation? *Philosophical Studies*, 173(1): 113–130. DOI: <https://doi.org/10.1007/s11098-014-0434-5>
- Hatano, G and Inagaki, K 1994 Young children's naive theory of biology. *Cognition*, 50(1–3): 171–188. DOI: [https://doi.org/10.1016/0010-0277\(94\)90027-2](https://doi.org/10.1016/0010-0277(94)90027-2)
- Hyman, J 2015 *Action, knowledge, and will*. New York: Oxford University Press.
- Löhrer, G and Sehon, S 2016 The Davidsonian challenge to the non-causalist. *American Philosophical Quarterly*, 53(1): 85–96.
- Lombrozo, T and Carey, S 2006 Functional explanation and the function of explanation. *Cognition*, 99(2): 167–204. DOI: <https://doi.org/10.1016/j.cognition.2004.12.009>
- Malle, B 1999 How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, 3(1): 23–48. DOI: [https://doi.org/10.1207/s15327957pspr0301\\_2](https://doi.org/10.1207/s15327957pspr0301_2)
- Malle, B 2004 *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.
- Malle, B, Knobe, J and Nelson, S 2007 Actor–observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, 9(4): 491–514.
- Mayr, E 2011 *Understanding human agency*. New York: Oxford University Press.
- Monroe, A and Malle, B 2017 Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *Journal of Experimental Psychology: General*, 146(1): 123–133. DOI: <https://doi.org/10.1037/xge0000234>
- Ryle, G 1949 *The concept of mind*. London: Hutchinson's University Library.
- Schult, C and Wellman, H 1997 Explaining human movements and actions: Children's understanding of the limits of psychological explanation. *Cognition*, 62(3): 291–324. DOI: <https://doi.org/10.1002/icd.548>
- Sehon, S 2005 *Teleological realism: Mind, agency, and explanation*. Cambridge, MA: Bradford Book/MIT Press.
- Sehon, S 2007 Goal-directed action and teleological explanation. In: Campbell, J K, O'Rourke, M and Silverstein, H S *Causation and explanation*. Cambridge, MA: MIT Press. pp. 155–170.
- Sehon, S 2010 Teleological explanation. In: O'Connor, T and Sandis, C *A companion to the philosophy of action*. Oxford: Wiley-Blackwell. pp. 121–128.
- Shapiro, S 1997 *Philosophy of mathematics: Structure and ontology*. New York: Oxford University Press.
- Skow, B 2018 *Causation, explanation, and the metaphysics of aspect*. Oxford: Oxford University Press.
- 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.

- Tanney, J 2013 Ryle's conceptual cartography. In: Reck, E H *The historical turn in analytic philosophy*. Basingstoke: Palgrave Macmillan.
- Thompson, M 2008 *Life and action: Elementary structures of practice and practical thought*. Cambridge, MA: Harvard University Press.