



Title	Two ways to understand causality in agency
Authors(s)	Stout, Rowland
Publication date	2007-11
Publication information	Stout, Rowland. "Two Ways to Understand Causality in Agency." De Gruyter, November 2007. https://doi.org/10.1515/9783110898798.137 .
Publisher	De Gruyter
Item record/more information	http://hdl.handle.net/10197/4959
Publisher's version (DOI)	10.1515/9783110898798.137

Downloaded 2026-05-01 23:41:42

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

Penultimate draft of Rowland Stout, “Two ways to understand causality in agency”, for
Anton Leist (ed.), Action in Context.

Section 1 – Introduction

I want to argue for a causal approach to agency based on an Aristotelian model of causal processes rather than on the more familiar model of a network of causal links.¹ Processes in this Aristotelian model are the realisations of potentialities. I will talk of mechanisms as the things that have potentialities; so processes can also be thought of as the workings of mechanisms.

An influential philosophical conception of our mind’s place in the world is as a site for the states and events that causally mediate the world we perceive and the world we affect. According to this conception, states and events in the world cause mental states and events in us through the process of perception. These mental states and events then go on to produce new states and events in the world through the process of action. Our role is as hosts for these states and events that causally mediate the states and events on the input side and those on the output side.

The picture can be made a bit more complicated by adding extra loops. So mental states and events interact with other mental states and events before they cause the body to move. And feedback loops should be added so that the effects of these bodily events

¹ I have been arguing for this in various ways, some of which recur here, for some time now. See Stout (1996; 2002; 2005; 2006).

themselves cause further mental states and events which then cause different bodily and worldly states and events. But however complicated the causal chain is – perhaps looking more like chain mail - the structure of states and events in the middle of the chain constitute the mind, and causally mediate non-mental inputs and non-mental outputs.

This picture can be described as Cartesian without yet committing it to any sort of immaterialism. It is Cartesian inasmuch as it takes the role of the person and their mind to be located in the middle of the chain of events passing from perceived inputs through the person's body to achieved outputs, just as Descartes in *The Passions of the Soul* placed mental events sandwiched in a causal chain between two sorts of movements of the pineal gland, the first affected by external stimuli and the second causing the body to move.

Many philosophers, often influenced by Wittgenstein, have rejected such a picture even without the immaterialism, thinking that in some problematic way it still makes mental states and events internal and unworldly. They also find it difficult in this picture to give reasons for action their proper role in agency.

How might reasons for doing things – teleological reasons in particular - be thought to have a place in action given the Cartesian model of our causal place in the world? First they might be taken to be facts about the world outside. The fact that the only way to get water is to dig a deep hole under this palm tree is a reason for digging the hole if you need or want to get water. But in the Cartesian picture this fact does not figure in action.

It figures in perception by causing various mental states including the belief that the only way to get water is to dig the hole. But once it has caused this belief that fact has no further causal role in the story. The belief combines with other mental states to cause the desire or intention to dig a hole and this desire or intention causes the body to move.

If we accept all this then we must deny that action essentially involves reasons. The problem simply put is that reasons figure on the input side of the Cartesian picture and action figures on the output side. The two sides exist independently of one another. So reasons turn out not to be essential to action. We might try to avoid this separation of reasons and action by taking reasons to belong to the output side of the picture. This would be done by describing the mental states and events (or facts about those mental state and events) as the reasons for action. So we could say that your reasons for digging the hole are your desire for water combined with your belief that digging the hole is the way to get water.

But this strategy undermines the idea of reasons for action. Facts about your mental states are not facts about means and ends. So they certainly cannot count as teleological reasons. And indeed as many people, including Fred Stoutland in this volume, argue, such facts are not really reasons for action.² At best they are reasons *why* you act. Someone's knowing reasons why you act can make your action rationally intelligible to them, but such reasons do not *favour* or *justify* your acting that way. It is the fact that digging a hole is the way to get water that favours or justifies digging a hole; the fact that you believe that digging a hole is the way to get water does not.

² See also Dancy (2000) and Stout (2004), as well as Stoutland (this volume).

It may seem that this Cartesian picture is quite inescapable if we acknowledge the causal nature of perception and action. How else could our mental states and events stand in the world, causally receptive and causally active, except as links in the causal chain (albeit a chain with feedback loops) from the world we perceive to the world we act on? With this worry in mind philosophers who reject the Cartesian picture very often tend to reject causal approaches to perception and action altogether. But I think that the apparent inescapability of construing our causal role as the role of links in a causal chain derives from a prevalent but quite avoidable conception of causation itself. When we see our causal role in another way then we can reject the Cartesian picture without being anti-causalists.

This prevalent conception of causation takes a basic causal process to be constituted from two particular events or perhaps complexes of events and states – the cause and the effect - and a generic relation of causation linking them. The causal world can be described according to this conception by specifying which structures of events and states are linked to one another by this relation. The causal world will look like a huge chain or network of events and states linked together. A diagram of such a causal world will have names for the events and states and a complex structure of arrows linking these names.

Let me call this the causal chain model of causation. The philosophy of causation is dominated by this model. The questions it asks concern how to understand the causal relation. For example, is it to be understood as a counterfactual relation linking two

events? To what extent does causation involve necessitation? Does it only link events or also states and perhaps facts or other things?

But this is certainly not the only model of causation. There is also an Aristotelian model that allows in addition to a cause and an effect not just a generic relation of causation linking them, but individual causal processes. In this model we can ask what the cause is and what the effect is, but in addition we can ask how the effect results from the cause – i.e. what the mechanism is. And this how-question is not to be answered just by introducing more links in a causal chain. It is answered by specifying the potentiality whose realisation is the process that the effect belongs to.

This Aristotelian model of the causal world cannot be fully pictured as a network of arrows, since each arrow would have to represent a different sort of causal process. It is a world of mechanisms having potentialities which are realised in the workings of the mechanisms and which realise each other. And applying this model to our minds' place in the world we can still take minds to stand between the perceived world and the affected world – causal intermediaries in a sense. But our mental states are not taken to be causal intermediaries in the sense of being nodes in the middle of a chain of causal links – caused by the inputs and causing the outputs. Instead they are taken to be the states of causal mechanisms transforming the inputs into outputs. According to this conception, states and events in the world are inputs into causal mechanisms that constitute our nature as both subjects and agents. Essential to these mechanisms are the potentialities whose realisations constitute the workings of these mechanisms. The way

our minds are is the way these mechanisms work – the way they transform perceived inputs into produced outputs. These states of mind are causally affected by the environment as well as by each other; but on this model a change in one's state of mind is not an essential part of a causal chain between the environment and one's behaviour.

There has been a revival of interest in mechanisms recently in the philosophy of science. See Machamer *et al* (2000), though it is possible that these philosophers mean something slightly different by mechanisms. An important earlier attempt to argue for an Aristotelian conception of causation is Harré and Madden (1975). And the task of trying to understand dispositions and potentialities has always been part of the remit of contemporary metaphysics (e.g. Mumford, 1998, working on well-established debates) even when this has not been taken to be part of the task of understanding causation. Thinking of the mind in terms of mechanisms, potentialities and dispositions is not new either. While Ryle (1949) is the most famous proponent of a dispositional view, Putnam's (1967) machine-table functionalism, taking mental states to be analogous to states of a computer programme, does something similar.

In the philosophy of action too, some philosophers are either using or feeling their way towards the idea that agency needs to be understood in terms of the operation of certain sorts of mechanisms. I will refer to some of these later. But the debate here has not been set out clearly yet as a debate between two alternative conceptions of causation. When it is set out in this way I think that much of the debate that is supposed to concern causalism

versus anti-causalism would be much better recast as concerning alternative conceptions of causality in agency.

Gilbert Ryle's apparent anti-causalism serves as a good example. He claims that appealing to dispositions or potentialities in explaining things is not to provide causal explanations:

There are two quite different senses in which an occurrence is said to be "explained". ... The first sense is the causal sense. To ask why the glass broke is to ask what caused it to break, and we explain, in this sense, the fracture of the glass when we report that a stone hit it. The "because" clause in the explanation reports an event, namely the event which stood to the fracture of the glass as cause to effect. But very equivalently we look for and get explanations of occurrences in another sense of "explanation". We ask why the glass shattered when struck by the stone and we get the answer that it was because the glass was brittle. (1949, 88)

And in the same way, according to Ryle, when explaining actions in terms of reasons we are appealing to dispositions and so not providing causal explanations. His anti-causalism is anti-*causal-chain*-causalism. Alfred Melden is the same:

Stating the motive is not offering a (Humean) causal explanation of the action.

The explanation does not refer us to some other event – the motive – which explains causally how the action comes to be. (1961, 102)

He attributes the causal chain model of causation to Hume and argues that this model does not apply in the explanation of action. But this leaves open the possibility that an Aristotelian model of causation might apply to the explanation of action. Elizabeth Anscombe (1957) does the same thing. She rejects the idea that explaining an action is done by describing the ‘mental cause’, and in so doing she appears to reject causalism. And many of the crop of contemporary anti-causalists can be described in the same way. They reject a particular version of causalism – one which perhaps identifies reasons for action with mental causes – without considering other ways that reasons for action may be involved in causal processes.

In this paper I want to present some considerations in favour of thinking of agency in terms of the Aristotelian causal process model rather than the Cartesian causal chain model. These considerations by themselves do not show the causal mechanism model to be correct. They just show what might be gained in the philosophy of action if we could show this model to be correct. In each case the traditional approach that looks at what causes what is seen as failing because what seems to be important is *how* the behaviour is caused, not just by what.

Section 2 – Making events happen and states obtain

My first task is to motivate a causal approach to agency. The standard causal theory of action, which identifies action with mental states or events causing physical behaviour, employs the causal chain conception of causation. My defence of causalism will not go through this route. So I want to start a bit further back with the claim that agency is best understood in terms of the intentional achieving of goals. Indeed we can go further back still and consider whether agency is best understood in terms of intentionally making events happen and states occur.

This is something which is certainly not accepted by everyone, and not even by everyone who accepts some sort of causal approach to agency – see Thalberg (1967). I want to consider a two stage argument for the claim that agency is best understood in terms of making things happen. Firstly, agency should be understood in terms of intentionally doing things. Then, doing things should be understood in terms of making the things done happen. For example, raising one's arm is making one's arm rise. So agency should be understood in terms of intentionally making things happen.

The first stage of this argument is not particularly controversial. Someone might object that failing to do something is just as much a manifestation of agency as doing something. But trying and failing to do something is still doing something. And in any

case, it is likely that failed trying is best understood in terms of successful trying. So I think we are safe to start thinking about agency in terms of intentionally doing things.

The second stage of the argument is much more tricky. It might be thought that action is a paradigmatic causal concept. Philosophers have often looked for accounts of natural causation in terms of our concept of human agency. But it would not follow that we should seek a causal analysis of doing, since there may be no causal notion more basic than that of doing by which doing should be understood. No one doubts that causation itself is a causal concept, but causation should not be understood in causal terms. If doing is a very basic causal concept then a causal analysis of doing might be as futile as a causal analysis of causation.

The example of raising one's arm is a very questionable choice as a paradigm of agency. Philosophers of action very often look at these very simple acts or achievements as the targets of their analysis. But it is not at all clear that what characterizes agency are acts like this rather than activities. Many of the things we do seem to be activities rather than achievements. Eating a banana, going for a walk, thinking about a problem, chatting with a neighbour are all doings that do not seem to amount to mere achievements. Eating a banana is not just achieving the goal of the banana being eaten. I can achieve that goal by getting my four year old daughter to eat a banana, without having to eat it myself. And it is not obvious that it should be understood in terms of making something happen or in causal terms at all.

The difficulty here is finding something other than the activity of eating the banana that counts as the thing caused when you eat a banana. It is certainly not right to identify eating a banana with causing oneself to eat a banana. For in that case eating a banana would be causing oneself to cause oneself to cause oneself and so on *ad infinitum* to eat a banana.

But is there anything other than the activity itself that might be the thing one causes when one acts? In the case of actions which are movements of one's body there is a very obvious candidate. My raising my arm might be identified with my causing my arm to rise. My arm rising is not an action – it is not what I do. To employ the Aristotelian terminology made familiar by Jennifer Hornsby's work, it is an intransitive movement, whereas moving my arm is a transitive movement (Hornsby, 1980, 2).

But even in this limiting sort of case of an action, it might be objected that I am not really causing my arm to rise; I am just raising it. The idea of causing my own arm to rise might sound rather odd – as if I have to employ some non-standard device like autohypnosis to make the arm rise. But I think this oddness may easily be attributed to the fact that in normal cases we have a simpler way to express what happened – namely that I raised my arm. The implication that there is some non-standard route to the arm rising when I make it rise can easily be cancelled. Suppose there was something else like a force field or an electric shock that might have made my arm rise, though it did not. Then a sensible question for someone to ask is: “Did you make that happen? And the sensible reply is: “Yes I caused my arm to rise.”

So I think it is right to identify moving one's body with causing one's body to move. But on the face of it not everything one does is moving one's body. Most of what one does involves moving one's body in some ways, but it is not at any rate obvious that eating a banana, going for a walk, or having a chat with one's neighbour should be *identified* with moving one's body, even in a very complex way. Now there are some influential arguments in favour of the non-obvious claim that anything one does is a case of moving one's body.³ But even if one rejected these arguments there is a more plausible fall-back position, which is that everything one does is a case of achieving some structure of intransitive results. These results might include one's body moving in certain ways. But they might also include things happening outside the body.

It is too simple to say that my eating a banana is my making the banana get eaten; after all I can make the banana get eaten without eating it. But perhaps it is not so far wrong to identify it with my making a structure of events/states happen/obtain that are characteristic of my banana eating – i.e. roughly speaking by making the banana become ingested through my mouth into my digestive system. Perhaps in a similar way my going for a walk is my causing some characteristic structure of achievements associated with going for a walk. It is not obvious exactly how to describe such a characteristic structure of achievements; but it looks like a perfectly reasonable philosophical task to work it out.

One consideration in favour of thinking of actions in terms of achievements is that with these activities it is always still possible to try but fail to do them. Although it would be

³ See Davidson (1980, essay 3).

strange it would not be impossible to try but fail to eat a banana, go for a walk or chat with one's neighbour. Perhaps you gag every time you attempt to swallow a bit of banana, the wind blows you over every time you take a step or the sound of traffic means that you cannot hear a word your neighbour is saying. This suggests that with these activities there is something that can be successfully achieved or not. And this in turn suggests that these activities involve the achievement of goals. If eating a banana is correctly described as achieving a particular structured goal then it is also correctly described as making that goal become achieved.

So I propose that a subject, S , φ -ing can be identified with S causing a structure of results characteristic of their φ -ing . Certainly this identification seems pretty awkward in some cases. But the move I make in the next section may help to reduce this awkwardness. For there I will argue that we can give a causal account that does not merely state what is caused, but states how it is caused. My eating a banana may then be seen to be not merely causing a structure of results characteristic of my eating a banana, but causing that structure of results by means of the operation of my eating mechanism. The thing about these activities is not just that they realise the capacity of agency as such, but also that they realise particular further instinctive or learnt capacities. This idea can be found a place in the Aristotelian model.

Section 3 – How results are caused

A traditional schema for analysing a causal notion – F – is as follows:

S F -s O if and only if some aspect of S causes O to have some quality characteristic of F .

Consider the wind drying the washing. Using the traditional schema this is to be understood as some aspect of the wind causing the washing to be dry. Which aspect of the wind should be specified here is up for discussion. Perhaps it is the wind itself that is doing the causing or perhaps it is some event or state associated with the wind. This traditional analysis explains the causal notion in terms of a causal link between two things – a characteristic cause and a characteristic effect. As such it uses what I have called the causal chain model of causation.

The obvious inadequacy of this account of the wind drying the washing is that by only specifying the cause and the effect it misses out the way the effect is caused, and this way too has to be characteristic of wind-drying. This can easily be seen by constructing a deviant causal chain counterexample. Suppose the wind blew open the window, which banged into the button that turned on the tumble dryer in which the washing soon dried. The wind caused the washing to be dry but it did not dry the washing. This is because it caused the washing to be dry in the wrong sort of way. It did not result in the washing becoming dry through the realisation of a wind-drying mechanism.

This explanation of what has gone wrong with the causal chain approach suggests an alternative way of analysing causal notions – what I have called the causal process model.

S F-s O if and only if *O* having the *F*-quality belongs to a process that is the working of an *F*-mechanism (the realisation of its *F*-potentiality) embodied by *S*.

For example, the wind dries the washing if and only if the washing being dry belongs to the working of a drying mechanism embodied by the wind. “*F*” appears on both the left and the right hand side of this biconditional, which means that this is not a reductive account. But there is no vicious circularity here. The characteristic results that go with a certain kind of causal notion can be established at least partially independently, and then we can establish on any particular occasion whether something belongs to a process that is the working of a mechanism with these characteristic results. For example, the characteristic results of the mechanism of the wind drying washing are something like this: the longer the washing is in the wind and the stronger the wind the dryer it becomes, until it is completely dry. If the drying of the washing belongs to a process that is the working of a mechanism with these characteristic results embodied by the wind, then we can say that the washing was dried by the wind.

In the deviant causal chain example the washing becoming dry does not belong to the working of such a mechanism. There is a drying mechanism in this example, but it is not

embodied by the wind. And the way the relevant bit of the mechanism that is embodied by the wind works does not have these characteristic results. In general the causal process model is immune from deviant causal chain counterexamples precisely because it does not employ the idea of a causal chain.

However this account does employ ideas that many modern philosophers of causation find deeply suspicious. First of all is the idea of a mechanism. A mechanism is identified by its potentiality. Here I am using the Aristotelian notion of a potentiality. Aristotle said: “Motion is the fulfilment of what exists potentially, in so far as it exists potentially” (*Physics*, 201a10-11). So a process is the realization of a capacity or disposition for certain results in certain circumstances. You have to characterize a structure of stages to specify the potentiality. But what is required for the process to be happening is not just that that structure of stages is in train, but that there is a potentiality for such a structure and that this potentiality is being realised.

There are two sides to something having a potentiality. There is the set of statements that describe what the potentiality is a potentiality for. And there is that underlying nature or set of conditions that grounds the potentiality.⁴ These latter conditions are divided into two categories – the conditions that constitute the potentiality or mechanism itself and the conditions that constitute the realisation of the potentiality or the operation of the mechanism. Usually this distinction is made in a pragmatic way, and for simplicity we can lump them all together as the underlying conditions of the process.

⁴ Philosophers opposed to this talk of mechanisms and potentialities are also suspicious of this idea of a grounding relation, and certainly the Aristotelian approach owes a treatment of it, even though I will not attempt one here.

Then there is the notion of belonging to a process. For the wind to dry the washing it is not sufficient that there is a process that is the working of a drying mechanism embodied by the wind and the washing becomes dry; the washing becoming dry must belong to that drying process. This leaves some important work to be done in the philosophy of causation in explaining what it is for an event to belong to a process. And until this work is done the suspicion that something like deviant causal chain counterexamples may be found cannot be completely eradicated. But process causation is certainly no more mysterious than event causation. To explain what it is for one event to result from another is at least as difficult as explaining what it is for an event to belong to a process.

Finally there is the notion of embodying a mechanism. For the wind to dry the washing not only must a drying mechanism be working and the washing becoming dry belong to that process, but also the wind must embody the drying mechanism. After all in the deviant causal chain example of the wind blowing the window into the tumble dryer which dries the washing, the washing becoming dry does indeed belong to a drying mechanism – just not a drying mechanism embodied by the wind. A full account would have to explain what it is for a thing to embody a mechanism or potentiality.

One further worry people have with this sort of account is that introducing potentialities and mechanisms provides empty explanations. Citing the dormitive virtue of opium has little explanatory value when trying to explain why opium puts you to sleep. Likewise, saying that the wind dries the washing if and only if the washing becoming dry belongs to

a process which is the realisation of a drying mechanism embodied by the wind does not have much explanatory power.

Now I do want this causal process account of causal notions to apply to the claim that opium put a patient to sleep. I want to say that it is true that the opium put the patient to sleep if and only if the patient falling asleep belonged to a process that is the realisation of a dormitive virtue or potentiality embodied by the opium. I also want to say that this claim is explanatory. But it does not explain why or how the opium put the patient to sleep; it explains what it means to say that the opium put the patient to sleep. So no wonder it has the sound of an empty tautology. Also it is not the end of the explanation but only the start. For one must go on to explain what a dormitive potentiality is – what the characteristic results of such a process are. And finally, although neither the claim that the opium put the patient to sleep nor what the causal process model takes to be the equivalent claim about the realisation of the dormitive potentiality explain how the opium put the patient to sleep, they both do explain why the patient fell asleep.

Let us go back to agency. In the previous section I argued that we should start by trying to understand what it is for an agent intentionally to make an event happen or a state obtain. More generally, we should start by understanding what it is for an agent intentionally to cause things to be *G*, where *G* stands for the intransitive characterisation of what they are achieving. A typical way for the causal chain model to explain this is the standard causal theory of action, which works in the following sort of way, though of course there are many variations of this:

An agent intentionally causes things to be *G* if and only if the agent's intending to cause things to be *G* causes things to be *G*.

The Aristotelian causal process model gives a different approach as follows:

An agent intentionally causes things to be *G* if and only if things being *G* belongs to a process that is the working of an intentionally-causing-things-to-be-*G* mechanism embodied by the agent.

This awkward phrase, “an intentionally-causing-things-to-be-*G* mechanism”, needs a bit of explanation. The point is that the way this mechanism works must be characteristic of intentionally causing things to be *G*. What the characteristic results are of an intentionally-causing-things-to-be-*G* mechanism is a central question of the philosophy of agency, and I will simply provide what I take to be the answer without any argument. I think the answer is that this involves sensitivity to reason – in particular teleological reasons. When we intentionally do things – cause things to be a certain way – we do them for reasons. And very often these reasons are teleological; we do things for the sake of other things.

For example, I walk to the letter box because that is what I should do in order to post the letter. A couple of qualifications are required. Firstly it is possible to act intentionally but not for any particular reason. I might intentionally eat a banana but not do it because

I was hungry or even because I particularly felt like it. I just chose to eat a banana and went on to eat it. But even if the action as a whole has no reason, the process of acting is still sensitive to reason. The top-level goal of eating a banana may just be given; but my behaviour is still sensitive to what I should do to achieve it.

Secondly it must be possible for quite irrational behaviour still to count as done for a reason. Somehow we must allow for behaviour that is rational within an irrational system. This behaviour may be irrational absolutely, but relative to that system it is rational. My suggestion for accommodating this is to require that the behaviour that accords with reason does so relative to a *version* of practical rationality. A version of practical rationality is to be construed as a way of making recommendations for action. It must involve some sort of means-end rationality. But it may also embed certain goals which need not be justified by this or any other version of practical rationality, as well as certain assumptions which need not be justified either and indeed may simply be false. For example the goal of eating a banana may be fixed in a version of practical rationality, and that version might also embed the quite false assumption that there is a banana in the kitchen. According to this version of practical rationality, and assuming that there is no better way to get a banana and so on, going to the kitchen is the thing to do. This goal of going to the kitchen is not fixed in that version of practical rationality but derived from applying it to the agent's environment. In general, a version of practical rationality, although it may embed certain fixed assumptions and goals must also involve some sort of sensitivity to the way things are.⁵

⁵ See Stout (1996 and 2006) for attempts to work this out in some detail.

So my proposal is that an intentionally-causing-things-to-be-*G* mechanism is simply a mechanism that characteristically results in what should happen for things to be *G*. It results in what should happen according to some particular, and possibly flawed, version of practical rationality. This gives us the following causal process theory of action:

An agent intentionally causes things to be *G* if and only if things being *G* belongs to a process that is the working of a mechanism embodied by the agent which characteristically results in what should happen according to some version of practical rationality for things to be *G*.

For another example, suppose an agent intentionally posts a letter. I argued in section 2 that this means the agent intentionally causes things to be characteristic of the general activity of their posting a letter. This will include, perhaps among other things, getting the letter into the postal system – i.e. causing the state of the letter being in the postal system to obtain. In the causal process model what it is for an agent intentionally to cause things to be characteristic of their posting a letter is for things being characteristic of their posting a letter to belong to a process that is the working of a mechanism that they embody that results in what should happen according to some version of practical rationality for such characteristic letter posting events/states to happen/obtain. When the way to get the letter posted (according to this version of practical rationality) is for the agent to walk to the post box the mechanism will result in the agent walking to the post box. If a mechanism with this sort of teleological sensitivity is what results in the letter

getting into the postal system then we can say that the agent has intentionally posted the letter.

An account of trying to achieve a goal, with or without actually achieving it, falls out naturally from the account:

An agent is trying to cause things to be *G* if and only if a process is happening that is the working of a mechanism embodied by the agent which generally results in what should happen according to some version of practical rationality for things to be *G*.

Trying involves the same teleological mechanism operating as acting does, but with trying there is no requirement that the achievement of the result belong to this process. So if the agent walks to the post box as a result of the working of the mechanism that results in what should happen according to some version of practical rationality for the letter to get into the postal system then they are trying to post the letter. If the agent tries and fails then they may not be doing what they should do in order to post the letter. But at least they are doing what they should do according to some possibly flawed version of practical rationality.

Section 4 – Deviance and Responsibility

For some time philosophers of action have been pushing their causal theories in this sort of direction. In 1964 Charles Taylor gave the following account of the teleological explanation of action:

To offer a teleological explanation of some event or class of events, e.g., the behaviour of some being, is, then, to account for it by laws in terms of which an event's occurring is held to be dependent on that event's being required for some end. To say that the behaviour of a given system should be explained in terms of purpose, then, is, in part, to make an assertion about the form of laws, or the type of laws which hold of the system. (1964, 9)

Adam Morton said: "intentional action is action that is guided by information to which it is responsive" (1975, 14). For many philosophers the problem of deviant causal chains has been what has motivated this sort of approach. Christopher Peacocke, responding to the problem of deviant causal chains, introduced the notion of differential explanation to attempt to make more precise the vague suggestion that "intentional behaviour is in some way characteristically sensitive to certain facts" (1979, 57). And David Lewis (1980)

provided a similar resolution to the analogous problem of veridical hallucination that faces causal accounts of perception.⁶

Converging on what I take to be the same target, philosophers of action have also introduced the idea of proximate or sustained causation.⁷ The rough idea here is that if there is any causal gap between cause and effect, then there is room for a deviant causal chain to be interpolated; so we should not allow there to be any causal gap between the characteristic cause of action and its behavioural effect. Process causation is sustained causation. The results of a process happen while that process is still happening. It is much less easy to see how a chain of causal links may sustain its results, since the earlier links are finished by the time the later ones come along.

Harry Frankfurt argued most explicitly for the claim I am making here that action must be regarded as a sensitive guided process, and thus that the causal chain theory of action is mistaken.

[T]he state of affairs while the movements [of a person's body] are occurring is far more pertinent [than the causes from which they originated]. What is not

⁶ Other philosophers who have introduced guided control into their accounts of action include Davis (1970, 23), Thalberg (1984), Audi (1993) and Bishop (1989).

⁷ See Brand (1984) and Thalberg (1984).

merely pertinent but decisive, indeed, is to consider whether or not the movements as they occur are under the person's guidance. It is this that determines whether he is performing an action. Moreover, the question of whether or not movements occur under a person's guidance is not a matter of their antecedents. Events are caused to occur by preceding states of affairs, but an event cannot be guided through the course of its occurrence at a temporal distance. (Frankfurt 1978, 158)

For Frankfurt, what distinguishes guided action is the causal mechanism not the causal antecedents.

Deviant causal chains are not a problem for this sort of account since action is not characterized in terms of a causal chain of any sort. There is no space in the process of a mechanism being realised in which a deviant causal chain can be interpolated. If a deviant causal chain were in place then the result would belong to the operation of a different mechanism – not a mechanism which works by making happen what should happen (according to some version of practical rationality) in order for some result to be achieved.

The situation gets a bit complicated if the deviant causal chain is itself a manifestation of some other agency. Robert Audi (1993) considers deviant causal chains due to alien intermediaries. The example is of Tom, who intends to look at his watch to shorten a

conversation. Ann, the alien intermediary, “likes to think she is making people do things that they would do anyway” (p. 164). She presses her buttons and Tom, as a result, looks at his watch when and in the way he was intending to. He does not notice a thing.

He is not intentionally looking at his watch; yet his behaviour does belong to the operation of a teleological mechanism – Ann’s mechanism.

The problem here is not whether to attribute agency to the behaviour, but to whom the agency should be attributed. This is similar to the deviant causal chain example for the drying washing where the washing becoming dry belongs to the working of a drying mechanism but not one that is embodied by the wind. In the account as I have stated it the agency belongs to whoever embodies the teleological potentiality the realisation of which results in the behaviour. And I have left this phrase “embodies the potentiality” unexplained. Although Tom’s body is part of the mechanism in this case, it is not his mechanism, but Ann’s. Certainly this needs to be worked out, but it is a challenge rather than an objection to the causal process approach to agency.

In just the same way, Fischer and Ravizza (1998) have developed a causal process theory of responsibility with two elements. Firstly for an agent to be responsible for some behaviour that behaviour must result from the operation of a “reasons-responsive” mechanism. Second, that mechanism must belong to the agent. Their motivation is not so much the problem of deviant causal chains but the related problem of alternative possibilities, something also dealt with by Harry Frankfurt, though some years earlier (1969). Frankfurt argued that responsibility for action does not depend on the truth of

some counterfactual conditional but rather on the nature of the actual process constituting the action.

We can adapt Audi's example above so that Ann only intervenes when Tom is not about to do what she wants him to do. Suppose that she has decided Tom will look at his watch whether he is inclined to do so or not. If he is not about to decide to look at his watch, she will press the levers in his brain and he will as a result decide to look at his watch. But if he decides to do this for himself she will do nothing. In this latter case, according to Frankfurt, we attribute responsibility to Tom, even though in some sense he could not have done otherwise. Fischer and Ravizza explain this by saying that in this latter case Tom's behaviour issues from a reasons-responsive mechanism that is Tom's own. What matters is the nature of the actual process resulting in Tom's behaviour, not the existence or otherwise of alternative possible processes resulting in different behaviour.

So Fischer and Ravizza are proposing something like a causal process model. But this proposal is treated as a fix within the philosophy of action to the particular issue of responsibility. They presume that we might have a notion of agency that does not employ this model which can then be beefed up by introducing the model to provide an account of really responsible agency. I suggest instead that once we start thinking of agency in causal terms at all the causal process model should be in place.

References

Anscombe, G. E. M. 1957, *Intention*, Oxford: Blackwell.

- Audi, R. 1993, *Action, Intention and Reason*, Ithaca NY: Cornell University Press.
- Bishop, J. 1989, *Natural Agency*, Cambridge: Cambridge University Press.
- Brand, M. 1984, *Intending and Acting*, Cambridge MA: MIT Press.
- Davidson, D. 1980, *Essays on Actions and Events*, Oxford: Oxford University Press.
- Davis, L. 1970, *Theory of Action*, Englewood Cliffs: Prentice-Hall.
- Fischer, J. and Ravizza, M, 1998, *Responsibility and Control*, Cambridge: Cambridge University Press.
- Frankfurt, H. 1969, "Alternate possibilities and moral responsibility", *Journal of Philosophy* 66, 829-39.
- Frankfurt, H. 1978, "The problem of action", *American Philosophical Quarterly* 15, 157-62.
- Harré, R. and Madden, E. 1975, *Causal Powers*, Oxford: Blackwell.
- Hornsby, J. 1980, *Actions*, London: Routledge.
- Lewis, D. 1973, "Causation," *Journal of Philosophy* 70, 556-67.
- Lewis, D. 1980, "Veridical Hallucination and Prosthetic Vision," *Australasian Journal of Philosophy* 58, 239-49.
- Machamer, P. et al. 2000, "Thinking about mechanism", *Philosophy of Science* 67, 1-25.
- Melden, A. 1961, *Free Action*, London: Routledge.
- Morton, A. 1975, "Because he thought he had insulted him," *Journal of Philosophy* 72, 5-15.
- Mumford, S. 1998, *Dispositions*, Oxford: Oxford University Press.
- Putnam, H. 1967, "Psychological Predicates", in Capitan, W.H. and Merrill, D.D. (eds.), *Art, Mind, and Religion*, Pittsburgh PA: Pittsburgh University Press, pp. 37-48.

- Ryle, G. 1949, *The Concept of Mind*, London:Hutchinson.
- Stout, R, 1996, *Things That Happen Because They Should*, Oxford: Oxford University Press.
- Stout, R. 2002, "The right structure for a causal theory of action", *Facta Philosophica* 4:1, 11-24.
- Stout, R. 2004, "Internalising practical reasons", *Proceedings of the Aristotelian Society* 44, 229-43.
- Stout, R. 2005, *Action*, Teddington: Acumen.
- Stout, R. 2006, *The Inner Life of a Rational Agent*, Edinburgh: Edinburgh University Press.
- Stoutland, F. 2007, "Reasons for Action and Psychological States", this volume, ??.
- Taylor, C. 1964, *The Explanation of Behaviour*, London: Routledge.
- Thalberg, I. 1967, "Do we cause our own actions?", *Analysis* 27, 196-201.
- Thalberg, I, 1984, "Do our intentions cause our intentional actions?" *American Philosophical Quarterly* 21, 249-60.