The right structure for a causal theory of action

In this paper, I argue that the way to understand causal notions is in terms of how something is caused rather than by what. To characterize how something is caused is to describe the way it is caused and this is to describe it as belonging to a certain process.

This idea of characterizing causal notions in terms of causal processes rather than causal antecedents is in itself not particularly new, but philosophers who have accepted this way of talking have usually spoiled the thought by working with an inadequate conception of a process. They have worked with a Russellian conception of a process as a chain of events rather than an Aristotelian conception of a process as the realization of a potentiality in a thing. This Russellian conception of a process is absolutely standard in work on philosophy of action from Davidson onwards. But characterizing something as belonging to a process in the sense of belonging to a causal chain of events is still to characterize it in terms of its causal antecedents.

The most familiar problem with analysing causal notions in terms of causal antecedents is that of deviant causal chains. This problem arises if there is the possibility of something being caused by the right sort of antecedents but in the wrong sort of way – i.e. because it belongs to the wrong sort of process. Treatments of deviant causal chains have increasingly come to recognize the need for some account of causal processes, by including the condition of sensitivity and requiring sustained causation. But such treatments have generally lost their nerve, and failing to see the power of this approach, add in extra conditions to meet what I argue in section 5 are bogus counterexamples.

I will use the causal notion of action as my central case, though nothing depends on this choice. The same considerations apply to perception, knowledge and any other causal notion. And I will defend Harry Frankfurt’s 1978 treatment of action, arguing that despite the recent discussion in the philosophy of action literature by, among others, Audi, Bishop and McCann, Frankfurt’s account, properly understood, is immune to deviant causal chain objections. In particular, there is no problem of ‘alien intermediaries’.

One’s first thought about how to analyse a causal notion might be in terms of a cause of a certain sort, an effect of a certain sort and a simple causal relation between the two. Such an account will have the form:

A if and only if C causes E.
Starting from here the next questions are how to characterize C, how to characterize E, and, given that such accounts never quite seem to work as they stand, what other conditions need to be added into the analysis. Further questions concern identification. Is there an identifiable particular picked out by the causal notion? If so, is it the cause, the effect, the two together, or perhaps some emergent entity that exists in virtue of the truth of the causal claim?

The problem of deviant (or wayward) causal chains affects any causal theory that has this sort of structure, whatever sort of thing the causes and effects are supposed to be. C may cause E in such a way that makes it accidental that such an effect happens. The right sort of thing may cause E, but may cause it in the wrong sort of way. This problem applies to causal theories of action, perception, knowledge and reference. It also applies to causal theories of breaking, heating, flooding, pollinating, tearing, etc. It faces any philosophical account that has the structure: A if and only if C causes E. The point is that a causal process is not completely characterized by characterizing the cause and characterizing the effect. There is also the way the effect is caused by the cause.

Consider the following causal theories:

- An impact breaks a glass if and only if there is an impact of the glass and there is a breakage of the glass, and the former causes the latter.

- The cooker heats the chicken if and only if there is activity of the cooker and there is an increase in temperature of the chicken, and the former causes the latter.

- The rain flooded the river if and only if there was rain and the water in the river overflowed its banks, and the former caused the latter.

- The insect pollinates the plant if and only if the insect engages in some pollinating behaviour and the plant is fertilized, and the former causes the latter.

Relatively little philosophical imagination is required to construct deviant causal chain counterexamples for any such account. Suppose a glass is fitted with a remote-control device that makes it implode at the press of a button by an operator, but otherwise is so robust that it can be thrown around the room quite happily. On a certain occasion the impact of the glass startled the remote-control operator so much that they accidentally pressed the button and the glass became broken. The impact did not break the glass, even though the causal requirement is met.

Or consider a downpour of rain causing the river to overflow its banks by stranded the lock keepers at the pub. The rain was not enough to increase the volume of water in the river substantially but only served to strand the lock keepers. They could not get back to the weir to regulate the flow of water properly, and as a result of this the river overflowed. The rain did not flood the river even though the causal requirement is again met.
In such cases of deviant causal chains the effect has the right sort of cause, but the wrong way of causing it. Davidson expressed the problem for the causal theory of action by saying that the “point is that not just any causal connection between rationalizing attitudes and a wanted effect suffices to guarantee that producing the wanted effect was intentional. The causal chain must follow the right sort of route.” And Davidson went on to despair of spelling out “a way in which attitudes must cause actions if they are to rationalize the action.”

But I will argue here that Davidson’s implicit identification here of the notion of a way of causing with that of a route or a causal chain is a mistake. Davidson was quite right to despair of spelling out the route that would characterize intentional action, but he was not right to despair of spelling out the way in which actions must be caused by their reasons.

In examples of deviant causal chains there are various stages in the causal process that are somehow illicit and other stages that should be there but are not. The lock keepers being stuck in the pub is not a proper stage in a process of a river being flooded by rain. A stage that should be in the process is that of the excess rain increasing the volume of water in the river. The activation of a remote-controlled implosion device is not a proper stage in the process of a glass being broken by an impact. So it is very tempting to try to fix a causal theory by specifying what intermediate stages there should and should not be in the causal chain. We might require that the causal chain in a process of flooding by rain should not pass through something other than the rain and the river. Or we might require that it should pass through this, that and the other event.

This would give us a theory of the form:

\[ A \text{ if and only if } C_1 \text{ caused } C_2 \text{ which caused } C_3 \ldots \text{ which caused } E, \]

combined with some extra conditions about what is not allowed to be in the chain. It is still a theory where the causal process is characterized by spelling out causes and effect, just more of them. And other conditions need to be included too.

This sort of response to the problem of deviant causal chains looks very unpromising. The inclusion of a set of extra negative conditions is ad hoc and means that we no longer have a simple causal account of a simple causal notion. Moreover these negative conditions invariably rule out things that should not be ruled out and include things that should be ruled out. In the case of the rain flooding the river, the lock keepers being stuck in the pub might have been a causal stage in the flooding process. The rain might have increased the volume of the water in the river significantly, while at the same time, if it had not also stranded the lock keepers in the pub, they would have been able to regulate the weir and avoid the flooding. But equally, it might not have been a causal stage in the flooding process.

Trying to shore up a causal theory by adding the requirement of certain intermediate stages is also a fairly hopeless strategy. Again, it leads to ad hoc and over-complicated

---

1 Davidson (1980, 79).
accounts. But the most obvious objection is that it will always be possible to interpolate a deviant causal chain between any two of the new stages. However close together the stages are it will always be possible to devise a situation in which some absurd and accidental causal process links them.

One response to the failure of a standard causal theory of some notion would be to deny that that notion is causal at all. But, although this strategy may work for some apparently causal notions, it is not a good strategy generally. Standard causal theories fail for every apparently causal notion, and we do not want to conclude that no notion is causal.

An alternative response would be to allow that these (or some of these) notions are causal but to deny that any causal theory is possible. The thought here might be Anscombe’s thought that the notion of a cause is less basic that the notions of breaking, flooding, pollinating, etc, and so these latter notions should not be analysed in terms of the former. Instead, causation should be understood to be an abstraction from these more primitive notions, while the notions of breaking, flooding, etc should be taken to be sui generis.

Certainly, the concept ‘cause’ seems to be a harder one to grasp than the concept ‘break’, and this suggests that it is not the more basic concept in terms of the progression of understanding. But this does not mean that there is no piece of philosophical theory that can be found to explain what all these causal notions have in common and why they all count as causal. Although it may turn out that there is no such general philosophical account of causal notions, the failure of the standard account does not warrant this conclusion. For there may be some room for improving the structure of the standard account. This possibility is what I explore next.

The problem with all the standard causal theories might be that they assume that the right way to characterize a causal process is by specifying the actual stages that the process does and does not go through. They assume a Russelian conception of processes, in which a process is identified with a causally connected structure of stages – a causal chain. Russell wrote: “Motion is the occupation by one entity of a continuous series of places at a continuous series of times”. This claim can be extended to processes generally so that we have the claim that a process is a continuous series of states of affairs. A process is a series of process-stages, perhaps satisfying some causal connectivity requirement. This position is captured by Wesley Salmon’s ‘at-at’ theory of processes, where the causal connectivity requirement is fleshed out in terms of the possibility of transmitting marks through the series of process-stages.

But there is an alternative conception of processes – one we owe to Aristotle. 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

---

2 Anscombe (1981, 137).
3 Russell (1903, section 442).
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 realized.

There are two sides to something having a potentiality. There is the set of conditional statements that describe what the potentiality is a potentiality for. And there is that underlying nature or set of conditions that grounds the potentiality. Knowing a potentiality is knowing a set of conditional statements, and this means that you are entitled to make inferences about what will be realized by that potentiality in different types of circumstance. Ryle called such conditional statements “inference tickets”. For example, in the case of a fragile glass, you know that if the glass is subject to an impact in a certain range then it will break. Knowing this, you are entitled to infer from the fact that the glass is subject to such an impact that it will break.

But there must also be something that entitles one to make the inferences that make up a potentiality. And this something is the nature and situation of the thing with the potentiality. In general, inferences may be justified by a number of other things: for example scientific method, authority, social convention, etc. But if any of these other things justifies the inference from describing a glass coming under an impact to describing its breaking we cannot thereby identify a fragility or a breaking potentiality. Only if it is the nature of the glass itself along with its environmental situation that justifies the inference can we attribute this potentiality to it.

The underlying structure that grounds a structure of inferences may be taken to be the conjunction of the existence of an underlying mechanism of a certain sort and the satisfaction of the operating conditions for that mechanism. But this division just raises unnecessary problems at this stage of the account. Suppose we see a car driving at 100 m.p.h. We might say that the car is the mechanism with the potentiality to go at 100 m.p.h. and having petrol in its tank is one of the operating conditions of that mechanism. On the other hand, we tend to think of the distributor on the carburettor as part of the mechanism rather than one of the operating conditions.

It is an interesting question why a distributor on the carburettor should be treated differently from the petrol in the tank. I can only assume that the answer must appeal to some pragmatic considerations about how easily each of these two things may be put in place. But for present purposes the issue should be ducked entirely. Let us take the underlying situation that grounds a process to be a working mechanism. So the existence of the mechanism and the satisfaction of its operating conditions are lumped together as the underlying conditions of the potentiality.

The process, which is the realization of a potentiality, is the same thing as the proper working of a mechanism. The process of a glass breaking under impact is the same thing

---

5 Ryle (1949, 121).
6 Ayers (1968, chapter 5) has an interesting discussion of this.
as the realization of the potentiality of the glass to break under impact, which is the same thing as the proper working of the mechanism of glass breaking under impact.

The potentiality is characterized by a set of general conditional statements, knowledge of which consists in entitlement to make a set of inferences. Such inferences are not universally valid. You are entitled to make the inferences only when their underlying conditions are fully in place. In order to know that a certain potentiality is being realized and so be entitled to make the respective inferences, you do not have to know exactly what these underlying conditions are; you just have to know that they are satisfied, whatever they are. In this way the entitlement to make such inferences is grounded in the nature of the situation.

Given this conception of a process, in order to characterize a causal process it is neither necessary nor sufficient to specify all the stages in the causal chain. Instead, one must describe the potentiality whose realization that process is. And to do this one must specify a structure of conditional statements.

This approach may look strangely like David Lewis’s (1973) non-Aristotelian approach to causation. In both cases counterfactual conditionals are crucial. The difference is that, for Lewis, conditionals are truths about the world of all possible worlds, whereas, according to the view being recommended here, the general conditional statements constitute descriptions of processes. The truth of the conditional claim is localized to the situation in which the underlying nature of the process is present.

So my suggestion is that the analysis of a causal notion should not be of the form:

\[ A \text{ if and only if } C \text{ causes } E. \]

Instead it should be of the form:

\[ A \text{ if and only if } E \text{ belongs to the realization of a potentiality which is described in terms of the characteristic sensitivity that goes with } A. \]

“\(A\)” 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 sensitivity that goes with a certain kind of thing 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 this sensitivity. For example, the law describing the mechanism of a glass breaking under impact will be something like this: if the impact on the glass is in such-and-such a range, then the glass will break. If the breaking of a glass belongs to a process that is the working of a mechanism with this sensitivity, then we can say that the glass broke under impact. A

---

7 Lewis (1973).
quite different law describes the mechanism responsible for the glass breaking as a result of a remote-controlled implosion.  

It is not sufficient that the conditional statements describing such sensitivity should merely hold in the situation. The breaking of the glass must belong to the process which is the realization of a potentiality that is described by those conditional statements. Put more simply, the process described by those conditional statements must result in the glass breaking. 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. But process causation is certainly no more mysterious that 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.

Consider now the example of the causal theory of action. The suggestion here is that we should not have an account of the form:

An agent is intentionally achieving E if and only if they intend to achieve E and their intending to achieve E causes the achievement of E.

Instead we should have an account of the form:

An agent is intentionally achieving E if and only if the achievement of E belongs to a process that is the realization of a potentiality which is described in terms of the sensitivity characteristic of intentional agency.

To characterize agency you do not just say that a certain sort of cause causes a certain sort of effect. You also have to say that the effect is caused in a certain sort of way. You have to describe the process of action in terms of what would have happened in alternative situations. You have to describe it in terms of the sensitivity of the process to different kinds of situations.

What is the sensitivity characteristic of intentional agency? One’s first thought might be that whatever is the intended goal it should be achieved. But describing agency in terms of achieving intended goals is to ignore the practical side of action. It is to ignore the possibility of acting intentionally while failing to achieve the intention. With a limited repertoire of basic acts at our disposal we do not simply achieve our intended goals when acting; we do what we should do in order to achieve our intended goals. The sensitivity characteristic of agency is teleological.

Sometimes doing what should be done in order to achieve a goal does not result in the achievement of the goal. What should be done in order to get some money might in the circumstances be to visit a cash point machine. But for one reason or another this may

---

8 However, if the remote-control operator is properly responsive to the occurrence of an impact on the glass and is committed to pressing the button when there is such an impact, then we can say that the impact breaks the glass even though the causal chain includes the remote-controlled implosion.
fail to result in actually getting the money. In such a case of acting with the intention of getting the money we say that the agent tried to get the money. Trying and failing is as much the realization of the potentiality to act intentionally as trying and succeeding.

So the suggested account of the causal notion of action is the following:

An agent is intentionally achieving E if and only if the achievement of E belongs to a process that is the realization of the potentiality to produce what should be achieved in order for the agent’s intentions to be achieved.

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.9

In 1975, Adam Morton said: “intentional action is action that is guided by information to which it is responsive”.10 In 1979, Christopher Peacocke 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”.11 And David Lewis provided a similar resolution to the analogous problem of veridical hallucination that faces causal accounts of perception.12

Converging on what I take to be the same target, philosophers of action have also introduced the idea of proximate or sustained causation.13 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. 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 structure of causal stages may sustain its results, since the earlier stages are finished by the time the later stages come along.

Harry Frankfurt in 1978 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

---

9 Taylor (1964, 9).
10 Morton (1975, 14).
11 Peacocke (1979, 57).
12 Lewis (1980). Other philosophers who have introduced guided control into their accounts of action are Davis (1970, 23), Thalberg (1984), Audi (1993) and Bishop (1989).
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 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.”  

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

Many of the other philosophers that I have listed have thought that this simple manoeuvre of switching from causal antecedents to causal processes or mechanisms in an account of action is insufficient. They have required something more to be included in their causal theories and as a result these theories have become baroque and implausible. But I argue that this is because they have failed to see the real strength of the idea.

Robert Audi has an approach to acting for a reason that builds on the notion of guided action. He says that “an action for a reason is one that is, in a special way, under the control of reason.” It is a response to, not a mere effect of, a reason”. But Audi sees a problem with this as it stands in the possibility of 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.

Audi’s response is to add an extra condition to his account ruling out alien intermediaries. But this move is entirely ad hoc. John Bishop is in a similar situation. He observes that alien intermediaries do not always stop a bit of behaviour from being intentional action. They may just count as sophisticated prosthetic devices for the agent. He argues that they only matter when they take over that part of a sensitive behavioural process that is responsible for the feedback loop. But Hugh McCann has argued successfully against this that an alien intermediary might allow Tom to do all the feedback bits of the process in his own brain and then step in helpfully to finish the job off for him. McCann’s conclusion is that all causal theories are doomed.

But I want to suggest instead that the problem is that Audi, Bishop and McCann are not seeing the real force of Audi’s Frankfurt-style account, which concentrates on special kinds of causal processes rather than special kinds of causal chains. Alien intermediaries are already ruled out, since the process resulting in Tom looking at his watch under Ann’s control cannot in fact be described in terms of the realization of a potentiality to produce what should be achieved for Tom’s intentions to be achieved. It looks like it can since

---

14 Frankfurt (1978, 158).
15 Audi (1993, 177).
17 McCann (1998, 123)
Ann is so respectful of Tom’s intentions. But the process really depends on what should be done given Ann’s intentions not Tom’s.

To see this, consider what would result from the Tom/Ann process if Tom intended to kill Ann. As it happens, Ann makes sure that Tom does not know that she exists. But to assess the counterfactual sensitivity of the process, we must consider what would result in this circumstance. Presumably the process working in this situation would not result in what should be done for Ann to be killed. Ann may be respectful of Tom’s intentions, but there is a limit. This shows that the process is not really sensitive to the rational requirements of Tom’s intentions, but to those of Ann’s intentions, which happen for the most part to coincide with Tom’s.

It might be objected that sensitivity to some condition cannot require that the appropriate result always occur given that condition. Mechanisms can go wrong or fail to operate. So could we say that the process results in what should be done in order for Tom’s intentions to be achieved but that the underlying conditions are not satisfied when the intended results are not in Ann’s interests? This would mean that the process, when it is happening, is sensitive to Tom’s intentions, but for some intentions the process can never be happening.

This is in fact an absurd suggestion. It is absurd because a set of inferences does not properly describe a process if any of those inferences fails to have any possible application. The Tom/Ann process that seems to result in Tom doing what should be done in order for his intentions to be achieved should properly be described as a process that results in what should be done for the achievement of those intentions of Tom’s that are not at odds with Ann’s interests. This is very different from describing a process as resulting simply in what should be done for Tom’s intentions to be achieved.

If a potentiality is properly described in terms of what should be done for the achievement of Tom’s intentions, then the circumstance in which Tom intends to kill Ann figures in the domain of types of circumstance of the function that describes this potentiality. I am requiring that it must be possible for the conditions underlying this potentiality to be present in that circumstance, and in such a circumstance what should be done for Ann to be killed will be done. But in the Audi example this is not possible, and so the process is not properly described as resulting in what should be done for the achievement of what Tom intends.

Of course, Ann may install herself as a slave for Tom, devoting herself entirely to making Tom do what he was going to do anyway. Suppose that she has no reasons for action that ever diverge or could diverge from Tom’s. However boring, horrible or self-destructive it is to make Tom do what he intends to do, she will stick it out forever. In this case I think it may be right to say that Tom is indeed acting intentionally and Ann is just a redundant prosthetic device – a pointless extra loop. Her own reasons have lost any independent causal role. In this case the process that results in Tom’s behaviour is truly sensitive to what should be done to achieve his intentions.
So thinking of action as the realization of a certain sort of potentiality rather than as the existence of certain sorts of causes and effects bypasses the problem of deviant causal chains. There is every reason to expect the same result for any other causal notion. For example, in the same spirit we might say that perception is the process that makes the way the world appears to a subject sensitive to the features of the world that are perceptually available to that subject. If a strange situation means that no such process is happening despite the fact that the accessible features of the world are causing the world to appear that way to a subject, then this is not a case of perception.

Thinking of action as the realization of a potentiality also helps with the problem of alternate possibilities. Even if Frankfurt-style counterexamples defeat the principle that in order to act freely it must have been possible to do otherwise, they do not touch the intuition that in order to act freely the action must belong to the realization of a potentiality to act in a number of different ways.\(^{18}\) The presence of the evil neuroscientist, ready to intervene at the first sign that I will not decide to do what she wants me to do may make it impossible for me to do otherwise. But the mechanism whose working results in my behaviour is one that, when working properly, would result in whatever should be done to achieve what I decide to do being done. The evil neuroscientist does not interfere with that mechanism, and so she does not interfere with my free action. Instead, she is in control of the situations in which that mechanism works properly and only in that sense controls what I do.

References


\(^{18}\) See Frankfurt (1969).