Uwe Meixner

The Case for Agent-Causation

There is a crack, a crack in everything.
That's how the light gets in.
Leonard Cohen

1 Symbols

event-causation: ➳ (understood as *sufficient or almost sufficient* event-causation, that is: causation that consists in the fact that event X is ne-cessarily, or almost necessarily, followed by another event Y).

narrow agent-causation: ⇓ (causation that consists in the fact that agent X makes an event Y actual, *without there being an event that causes* Y).[1]

(actual) physical event: ✷

(actual) agent: ■

X has experience Y: X → Y, or: Y ← X, or: X →→ Y, or: Y ←← X (X, in contrast to Y, is not temporally located; X has Y at the time represented by the tip of the arrow – or by the time that is represented by tip of the *second* arrow)

2 Conventions

If the symbol of an event is further to the left on the page than the symbol of another event, then this is taken to signify that the first event occurs earlier than the second.

"X causes Y" is taken to be a synonym of "X is a sufficient or almost sufficient cause of Y".

1 Concerning the phrase "X causes Y", see the second convention in Sect. 2 below.

The complete lack of ➳-*arrows* (black, horizontal arrows) *pointing* to an item signifies *lack of being* ➳-*caused* (event-causally caused); but the complete lack of ➳-*arrows pointing from* an item does not necessarily signify *lack of* ➳-*causing* (event-causal causing).

The complete lack of ⇓-*arrows* (white, vertical arrows) *pointing* to an item signifies *lack of being* ⇓-*caused* (agent-causally caused); but the complete lack of ⇓-*arrows pointing from* an item does not necessarily signify *lack of* ⇓-*causing* (agent-causal causing).

3 A picture of the causation of physical events as viewed in 19th-century physics

∞...➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳...∞
∞...➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳...∞
∞...➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳...∞
∞...➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳...∞

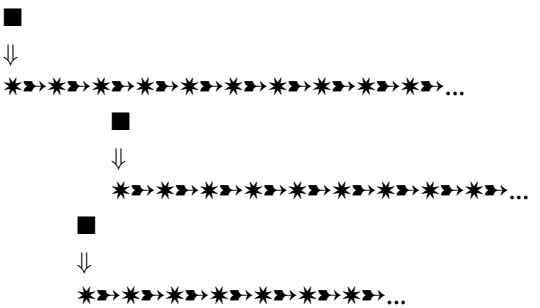
4 Three alternative pictures of the causation of physical events as viewed in 20th-century physics

Picture 1:

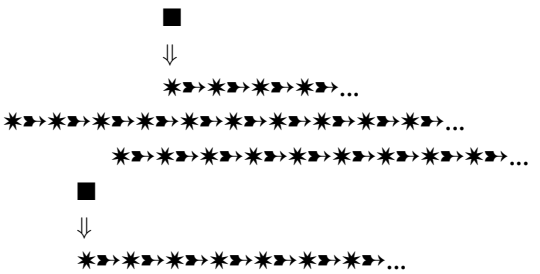
✷➳✷➳✷➳✷➳...
✷➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳...
✷➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳✷➳...
✷➳✷➳✷➳✷➳✷➳✷➳✷➳...

Picture 2:

■
⇓
✷➳✷➳✷➳✷➳...



Picture 3:

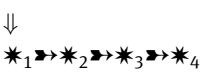


5 Central agent-causal concepts

Narrow agent-causation (of a physical event):

Note that narrow agent-causation excludes the "simultaneous" event-causation of the agent-caused event (as is suggested by the picture).

Broad agent-causation (of physical events):



Agent-causation (*simpliciter*) is here identified with *broad agent-causation*; it can be defined on the basis of *narrow agent-causation* and *event-causation*:

X (*simpliciter*) *agent-causes* Y $\stackrel{\text{df}}{=}$ X narrowly agent-causes Y, or X narrowly agent-causes some event Z and a chain of event-causation runs from Z to Y.

If an event Y is agent-caused (as defined) and event-causally traceable to, or identical to, *exactly one* event Z that is narrowly agent-caused, then *the time of the agent-causation of* Y is the time of (the occurrence of) Z.

It is assumed that each agent-caused event is event-causally traceable to, or identical to, *exactly one* narrowly agent-caused event. This is a plausible, yet a contingent assumption: there are unusual possible worlds in which that assumption is not true. But, plausibly, the actual world is not among those possible worlds.

An event that is agent-caused by X is an *action* of X.

An event that is narrowly agent-caused by X is a *direct action of* X.

An event that is agent-caused by X, but not narrowly agent-caused, is an *indirect action of* X.

An event that is agent-caused by X and that after its time of agent-causation is not experienced by X as being willed by X is an *implicit action of* X.

An event that is agent-caused by X and that after its time of agent-causation is experienced by X as being willed[2] by X is an *explicit action of* X.

Comment: It is obvious that the above-defined concept of action (see, above, the first definition) is a very wide, a very inclusive concept of action. For example, it allows events to be *actions* of X that are far removed, temporally and spatially, from X, or rather, from the relevant direct action (of X): the event which is the

2 The willing that is experienced with respect to an explicit action has the phenomenology of "imperativeness"; it also has the phenomenology of personal effectiveness. It is utterly different from *wishing*.

"original action" that starts an event-causal chain.[3] It does seem natural to add further necessary conditions for being an action (for example, the condition that only macroscopic bodily movements – behaviours – of X are actions of X).[4] I have nothing in principle against replacing the concept of action I propose by a more demanding concept. What I insist on, however, is this: *being agent-caused is a necessary condition for being an action*. Thus, the above-defined concept of action is the *minimal* concept of action.

6 A sufficient and agent-causal analysis of an action – of raising one's arm

That is: an analysis that provides a truly sufficient condition for being an action (in the particular case considered), and which is agent-causal in nature.

Wittgenstein in the *Philosophical Investigations* (Wittgenstein (2009), § 621):

> Let us not forget this: when 'I raise my arm', my arm goes up. And the problem arises: what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm? ((Are the kinaesthetic sensations my willing?))

There is a perfect solution to Wittgenstein's problem.

But first a

---

preliminary remark: In what follows P is *the subject of consciousness and agency* of a (let's assume: female) human being. Phenomenologically speaking, every one of us is the subject of consciousness and agency of a human being – a human being that encompasses that subject but is not identical to it. One cannot identify the subject of consciousness and agency of a human being with the human being in its entirety, since every one of us can truthfully say the following: "My eyes are much closer to me than my feet". It is to a large extent unproblematic – and common linguistic practice – to ascribe parts and properties of the entire human being *analogically* also to the human being's subject of consciousness and agency.

---

3 *What* is – relative to a direct action α of an agent – the spatial, temporal, spatiotemporal distance δ such that any event that is further away than δ from α cannot count as an action of that agent, although it be connected to α by an event-causal chain? (Cf. the second illustration in Sect. 10, (a).)

4 This may, in fact, be too strict a requirement.

Now, it is often true that someone's arm goes up. Consider a particular case: the arm of P goes up between t_1 and t_2.

The going-up of the arm of P between t_1 and t_2 is a physical event, call it "**R**". **R** is an event that *actually happens*; it is an *actual event* (not a merely possible event). (But actuality is not an intrinsic trait of events.)

P *raises her arm* between t_1 and t_2 if, and only if, **R** is (not merely an actual event but also) an *action* of P.
In other words (according to the definition of *action*):
P *raises her arm* between t_1 and t_2 if, and only if, P *agent-causes* **R** (which entails – but is not entailed by – the fact that **R** is an actual event).

In fact, if P is a normal human subject in a normal situation and raises her arm between t_1 and t_2, then **R** is an *explicit indirect action of* P, in other words: (1) P agent-causes, but does not narrowly agent-cause, **R**, and (2) P experiences **R** as being willed by her after the time of the agent-causation of **R**.

7 Nine insufficient analyses of raising one's arm

It is *not* (*conceptually*) *sufficient* for the fact that **R** *is an action of* P (i.e., that P *raises her arm* between t_1 and t_2) that **R** (the going up of P's arm between t_1 and t_2) is an actual event.

It is *not sufficient* for the fact that **R** *is an action of* P that **R** is an actual event and that P experiences **R** before it happens as being willed by her (i.e., that P experiences that she is – imperatively [cf. footnote 2] – willing **R** to happen).

It is *not sufficient* for the fact that **R** *is an action of* P that P experiences **R** before it comes about as being willed by her and that *this experience* event-causes **R** (via an event-causal chain).

It is *not sufficient* for the fact that **R** *is an action of* P that there is a physical event in the brain of P that event-causes **R** (via an event-causal chain).

It is *not sufficient* for the fact that **R** *is an action of* P that there is a *physically causeless* physical event in the brain of P that event-causes **R**.

It is *not sufficient* for the fact that **R** *is an action of* P that there is a *causeless* physical event in the brain of P that event-causes **R**.

It is *not sufficient* for the fact that **R** *is an action of* P that some event that intrinsically involves P, or is in any other non-causal way related to P, event-causes **R**.

It is *not sufficient* for the fact that **R** *is an action of* P that the combined desires and beliefs of P event-cause **R**.

It is *not sufficient* for the fact that **R** *is an action of* P that **R** is a rational (mediate or ultimate) goal with respect to the combined desires and beliefs of P and that the combined desires and beliefs of P event-cause **R**.

8 Deviant causal chains?

The agent-causal account of action is not subject to *the problem of deviant causal chains*, which besets the standard purely event-causal theory of action, forcing its proponents to accept as actions events that, intuitively, one does not want to accept as actions:

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally. (Davidson (1980), 79)

As long as the climber's loosening his hold is not agent-caused by him, it is not an action of his (not even if it is caused – in a deviant, non-normal way – by the climber's belief and desire). And Davidson's scenario indicates that it is precisely the case that the climber *does not* agent-cause his loosening his hold.

9 Causal responsibility and moral responsibility

Let X be an agent and Y a certain event:
X is causally responsible for Y iff Y is an action of X [*that is*: iff Y is agent-caused by X].
X is morally responsible for Y iff X is causally responsible for Y *and* … [here further conditions need to be added].

Principle of responsibility
A person is morally responsible only for those events that she is causally responsible for.

Query: But often a person is morally responsible for an event that she is not causally responsible for. For example, for an event that she could have prevented, but did not.

Answer: The best reaction to this query is to widen the concept of causal responsibility. Above we have what is in fact only the definition of *narrow* causal responsibility. A *wider* concept of causal responsibility – which allows to speak of causal responsibility also in cases where one cannot speak of it *so far* – can be defined in the following way:

X is causally responsible* for Y iff *either* Y is an action of X, *or* Y is an actual event without being an action of X, but some event Z which is *incompatible* with Y would have been an action of X if only X had willed Z.

The definiens of this definition is taken to be entailed by the state of affairs that X could have prevented Y, but did not.

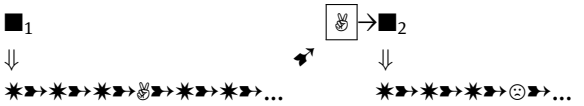
Replace "causally responsible" in the *Principle of responsibility* by "causally responsible*".

10 Illustrations

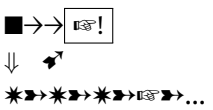
Let me illustrate the value of the concept of agent-causation for action analysis a little bit further:

(a) Two "chains of command" (one fairly short, the other somewhat longer):

(b) Combined "chains of command":



(c) The agent-causal interpretation of the result of the Libet-experiment:



11 Seven objections to agent-causation *answered*

Objection 1: Each instance of agent-causation must be either a causing of a physical event or of a non-physical event. Since the existence of non-physical events is rather questionable, agent-causation of non-physical events is rather questionable, too. Therefore, agent-causation can be a philosophically interesting option only if at least *some physical events* are agent-caused. But every physical event is already caused by a physical event. Hence: if one assumes that some physical events are agent-caused, one is making a superfluous assumption. There is, therefore, no good reason to assume the existence of agent-causation.
Response to Objection 1: Leaving entirely aside the question of the existence of non-physical events and their causation, it is still very likely true that *some physical events* are not caused (sufficiently or almost sufficiently) by any physical event. This much is strongly suggested by modern physics. There is, therefore, room for agent-causation; for agents may cause (may make actual) some of the physical events that are not caused by any physical event.

Objection 2: If there are in fact some physical events that are not caused – that is, not caused sufficiently or almost sufficiently – by any physical event, then the only reasonable conclusion regarding such events is this: that they – all of them – are not caused by anything at all, in other words: that their coming about is *to a significant extent* mere chance.
Response to Objection 2: Why is this supposed to be *the only reasonable conclusion*? No sufficient reason for this supposition is apparent. In fact, the very reasonable Principle of Sufficient Cause requires that *all* physical events that are not caused by any physical event still have *some sufficient cause*. And even if we may

not wish to appeal to the Principle of Sufficient Cause, lest we seem dogmatic, it also seems dogmatic to assert that all physical events that are not caused by any physical event have no sufficient cause. How could one be justified in *excluding* that *at least some* of those events *have* a sufficient cause?

Objection 3: But if some physical event that is not caused by any physical event had a sufficient cause, then this would constitute a violation of the Principle of Causal Closure of the Physical World, whether one identifies that principle with the proposition that *every sufficient cause of a physical event is itself a physical event*, or with the logically weaker proposition that *every physical event that has a sufficient cause also has a physical event as sufficient cause*. The Principle of Causal Closure of the Physical World, especially its logically weaker version, just cannot be violated.
Response to Objection 3: Wouldn't you agree, *on reflection*, that this, your last assertion, is a piece of sheer metaphysical dogmatism?

Objection 4: Well, not just metaphysics, also physics itself speaks against sufficient causes of physical events that are not caused by any physical event. For if some physical event that is not caused by any physical event had a sufficient cause, then the conservation principles of physics would be violated, because causation must manifest itself in the physical world by a change of energy and momentum.
Response to Objection 4: It is true that causation must manifest itself in the physical world by a change of energy and momentum. But if every such change violated the conservation laws, then these laws would be violated *all the time*, since changes of energy and momentum occur *all the time*. What is true is this: *presupposing* that the physical world is a physically closed system – that is, a system exchanging no energy or momentum with an outside –, the conservation laws forbid changes of energy and momentum which are such that they involve an increase or decrease in the sum total of energy, respectively, momentum; the principle of the conservation of momentum also forbids changes of momentum which are such that they involve a modification *in the total direction* of momentum. But it cannot be established – at least not given today's physics – that causing a physical event that is not caused by any physical event *must* involve a modification in the sum total of energy, or of momentum, or a modification in the total direction of momentum. One need not invoke here a physics-compatible causation of physical events by non-physical events; one can do better than this. Causing an event that is not caused by any physical event is best regarded as consisting *in the active resolution of a situation of physical indetermination*, that is, in choosing and actualizing *one* of several possible but, on the basis of the laws of physics, incompatible physical

events, each of which is compatible, on the basis of the laws of physics, with the given physical past. That situations of physical indetermination occur is allowed by present-day physics. And the combination of choosing and actualizing which resolves such a situation of physical indetermination must be regarded as the active work of an *agent*. This *causal work* does *certainly not* hurt the conservation laws, as little as the resolution of a situation of physical indetermination by sheer chance hurts those laws (as is admitted on all sides).

Objection 5: If this is what agent-causation consists in – narrow agent-causation, I suppose, according to your conceptual scheme –, then agent-causation does not seem to differ from fixed predetermination or, alternatively, from the workings of chance. Let me explain, along the lines of a thought-experiment which is due to Peter van Inwagen.[5] Let the situation of physical indetermination recur in exactly the same way an indefinite number of times. If the agent always does the same thing (chooses and actualizes the same physical event), then the agent appears to be predetermined. If, however, the agent does not always do the same thing, then the agent appears to be subject to chance, and subject to chance in the highest degree if the agent actualizes each possibility that is open to her in the (recurring, and recurring) situation of indetermination with the same frequency. How can you assure, as you surely wish to, that the agent is neither subject to predetermination nor to chance?

Response to Objection 5: If it is always the same event which is chosen and actualized in each recurrence of the same situation of physical indetermination, then this does not necessarily mean that the agent is somehow predetermined to choose and actualize *that* event. It may simply mean that rationality always tells the agent that she ought to actualize that particular event, and that the agent always decides to follow this constant advice of rationality. If it is *not* always the same event which is chosen and actualized in each recurrence of the same situation of physical indetermination, then this does not necessarily mean that the agent is subject to chance. It may simply mean that the agent in one recurrence of the situation decides to follow the constant advice of rationality to actualize a certain event, but in another recurrence decides to be irrational and to choose and actualize quite another event. I concede that, in an analysis of action that is based on agent-causation, there is in the causal explanation of action no going beyond the decision of the agent, *that is*: no going beyond the agent's initiating step of causation. In an analysis of action that is based on agent-causation, the causal explanation of action must stop with the agent's initiating step of causation. *But*

5 Van Inwagen (2002), 175–177.

this, in itself, does certainly not make the agent subject to chance. The agent is *not* subject to chance in her decision if, in her decision, instead of throwing dice she is following the advice of rationality (which, properly speaking, is *her own advice*: the advice that she is giving herself from the rational point of view). In turn, the agent's following the advice of rationality is, in itself, certainly not an instance of predetermination: rationality – since it is essentially *normative*, yielding, at best, only *ought to* and *ought not to* (relative to the agent's beliefs and desires) – is not a sort of prolongation of event-causal determination.

Objection 6: In order to be relevant for the physical world, agent-causation requires situations of physical indetermination. But while such situations may perhaps occur in the physical micro-world, they certainly do not occur in the human sphere. Agent-causation, therefore, is irrelevant for the analysis of physical human action.
Response to Objection 6: Suppose you were right and situations of physical indetermination did not occur in the human sphere. Then, indeed, agent-causation would be irrelevant for the analysis of physical human action – for the simple reason that there would be no physical human actions. There would be plenty of human outward behaviour, of course, but no physical human actions. No part of the physical history of the world would be made *by us*; the parts of that history in which we are involved would be made merely *through us*, with our assent or without, and very likely we would not even have a choice regarding assenting or not assenting to what is going on physically. You assert it as a certainty that there are no situations of physical indetermination in the human sphere. I do not believe that this is a certainty; I believe it is far from a certainty. But I am ready to admit: *perhaps* there are indeed no situations of physical indetermination in the human sphere. If this turns out to be true, then, I submit, philosophers should quit playing around with words; then they should have the intellectual honesty to *admit* that there are no human physical actions and that therefore nobody is truly responsible for anything in the physical world.

Objection 7: Well, that seems a bit panicky. Let's not panic here. I, in any case, won't panic. Though my entire physical behaviour is executed by a deterministic automaton, as I firmly believe, parts of that behaviour are actions of mine, for which I am indeed *truly*, *truly* responsible. I could not be more truly responsible for them: because they agree with what I consider, after careful deliberation, to be my most important goals and needs; because they agree with my essence, so to speak. I am free, see. I could not be any freer. – But be that as it may, returning to agent-causation, I would finally like to point out that agent-causation is a completely obscure idea. I have no idea what you mean when you say that an agent

causes a physical event *without* an event causing it, in other words, when you allegedly refer to an instance of *narrow agent-causation*, as you call it.
Response to Objection 7: The heart of causation is that the cause *makes a possible event actual*. Concepts of causation must at least *come near* to this idea. If a concept called "causation of this or that type" does not come near to that idea, then that concept has nothing to do with causation *conceptually*, but pays mere lip service to it. Having considered the various concepts of event-causation, I do not believe that any extant concept of event-causation comes nearer to *the making actual of possible events* than narrow agent-causation does. If you are telling me that you do not understand narrow agent-causation, you are in effect telling me that you do not understand *the making actual of possible events*. But *then*, how can you understand *any* proper concept of causation, *any* concept of causation that does not pay mere lip service to causation? I concede that in narrow agent-causation we are confronted with just the agent-causal relation, and that there is nothing that fits between the agent-cause and its effect. Event-causation does not have this kind of immediacy: it is founded on laws and mechanisms, and usually in event-causation, an event-causal chain fits in between cause and effect. All of this creates the illusion that event-causation is better understood than narrow agent-causation. But in fact, to the extent that event-causation *really* deserves the name "causation", it is no better understood than narrow agent-causation.

12 Agent-causation and freedom of the will

In his influential book *Das Handwerk der Freiheit*, the Swiss philosopher Peter Bieri writes that the freedom of the will consists in the will being determined in a rather specific way: by our thinking and judging.[6] Many philosophers find this definition of the freedom of the will entirely satisfactory, and all the more so because it is compatible with determinism. I do not believe that Bieri's idea of the freedom of the will is correct. Would my will be still free *if*, indeed, it were sometimes determined by my thinking and judging, but my thinking and judging, in turn, were always determined by causes that have nothing to do with my thinking and judging? Contrary to what Bieri and other compatibilist philosophers are satisfied to believe, I do not think that my will would be free under the condition just envisaged. But I concede that that condition comprises all that *my rationality*

6 Bieri (2001), 80: "Die Freiheit des Willens liegt darin, dass er auf ganz bestimmte Weise bedingt ist: durch unser Denken und Urteilen".

could at most amount to if human beings were in fact *deterministic automata*, as many brain-scientists believe and do not hesitate to proclaim with the full weight of their presumed scientific authority. A fortiori, that condition comprises all that my rationality could at most amount to *if determinism ruled the world*. If I am a deterministic automaton, which under determinism I must be, then my rationality could at most amount to this: though my thinking and judging are always determined by causes that have nothing to do with my thinking and judging, my will is sometimes determined by my thinking and judging. – Well, great. Am I then *truly* rational, even free? I do not think so. How, in the world, could my will, if it is ultimately determined by external factors that have nothing to do with my thinking or judging, be *truly* rational or free?

But perhaps human beings are not deterministic automata. There is room for reasonable doubt. And if human beings are not deterministic automata, then there is room for agent-causation, and room for *free will*, properly speaking. Specifically, an indirect explicit physical action Y of mine, event-causally traceable to a direct implicit physical action Z of mine, is *certain* to be *an act of my free* will if all of the following conditions are fulfilled:

(i) no agent other than me causes Z;
(ii) my causing of Z is not somehow determined, and it resolves a situation of physical indetermination;
(iii) Y is rationally intended by me *at least from* the time shortly before my agent-causing Y (= the time of my agent-causing Z) *to* the time shortly after my coming to experience that Y is willed by me, and
(iv) Y would not have happened without Z having happened.

13 The way from freely willed behaviour to agent-causation

1. An arm rises: the event Y takes place, and the subject of the human being whose (right) arm rose, P, declares that she freely raised her arm.
2. Suppose she is right. Then, *either* Y has an event-cause (2.**1**), *or* Y has no event-cause (2.**2**).

3. I opt for 2.**1**. Then, *either* there is a *complete and stopping* chain of event-causes[7] for Y (3.**1**), or there is none (3.**2**).
4. I opt for 3.**1**. Then, *either* there is a complete and stopping chain of event-causes for Y which is such that the first event of it is temporally close to Y and located in the brain of P (4.**1**), *or* there is no such chain (4.**2**).
5. I opt for 4.**1**. Then, *either* there is *exactly one* complete and stopping chain of event-causes for Y which is such as described in 4. (5.**1**), *or* there is not exactly one such chain (5.**2**).
6. I opt for 5.**1**. Then, let Φ be *the* complete and stopping chain of event-causes for Y which is such that the first element of it, $1(\Phi)$, is temporally close to Y and located in the brain of P; *either* there is a sufficient cause of $1(\Phi)$ (6.**1**), *or* there is none (6.**2**).
7. I opt for 6.**1**. Then, *either* there is exactly one sufficient cause of $1(\Phi)$ (7.**1**), *or* there is not exactly one such cause of $1(\Phi)$ (7.**2**).
8. I opt for 7.**1**. Then, *either* the sufficient cause of $1(\Phi)$ is P (8.**1**), *or* not (8.**2**).
9. I opt for 8.**1**.

Bibliography

Bieri, P. (2001), *Das Handwerk der Freiheit*, München: Hanser.

Davidson, D. (1980), "Freedom to Act", in D. Davidson, *Essays on Actions and Events*, Oxford: Oxford University Press, 63–81.

Van Inwagen, P. (2002), "Free Will Remains a Mystery", in *The Oxford Handbook of Free Will*, edited by R. Kane, Oxford: OUP, 158–177.

Wittgenstein, L. (2009), *Philosophical Investigations*, the German text with an English translation by G. E. M. Anscombe, P. M. S. Hacker and J. Schulte, revised 4th edition by P. M. S. Hacker and J. Schulte, Oxford: Blackwell.

7 A chain of event-causes for Y is *complete* if, and only if, it can neither be prolonged nor filled up. A chain of event-causes for Y is *stopping* if, and only if, it has a first element.