

Georg Gasser

Agent-causation and Its Place in Nature

1 Doubting Humeanism

Many contemporary philosophers are consciously or unconsciously in the grip of a Humean conception of reality. A Humean conception of reality claims that all that exists, ultimately speaking, are basic entities of a specific type (events, for instance) and relations between them. Relations are described as causal if their relata succeed temporally, are contiguous, are qualitatively similar and follow a repetitive pattern. For causation to happen, no further ontological ingredient is required. One can dub such a conception of reality “structuralist” and “actualist”: “structuralist” because the world is ultimately a complex structure of spatiotemporal relations depending on the specific distribution of basic entities, and “actualist” because the basic entities are locally instantiated qualities. Thus, to use Hume’s famous phrase, all ultimate entities seem entirely loose and separate; they seem conjoined but never connected. (Hume (1999), 72.26.) A prominent contemporary defender of this view is David Lewis. According to him,

all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another. (Lewis (1999), ix)

Such a view is, ontologically speaking, extremely sparse, which makes it attractive. It complies with the researcher’s old dream to explain nature’s complexity in strikingly simple terms. Take, for instance, the observation that salt dissolves in water. Explaining this observation requires no recourse to a substance, salt, or its special dispositional properties, such as being water-soluble; to account for the electrostatic process of hydration, all that we need to do point to the physico-chemical properties of salt and water.

Among the major motives of philosophers defending a Humean account is the worry that reference to dispositions and powers unnecessarily conflates our ontology and, even worse, opens the door to entities which science cannot grasp. The vase’s disposition to break appears invisible; all that we can see is that the vase breaks when struck with a hammer. Salt’s disposition to dissolve in water is unobservable; all that we can observe is that the grains of salt have dissolved after being placed in a sufficient quantity of unsaturated water. Given that the physico-chemical structures of the entities involved in these causal processes

Georg Gasser: Innsbruck University

are open for empirical investigation, why suppose that there are dispositional properties? Lewis emphasises this view when he writes that

[t]he point of defending Humean Supervenience is [...] to resist philosophical arguments that there are more things in heaven and earth than physics has dreamt of. (Lewis (1999), p. 474)

Without going into further detail, it should be obvious that a conception of the world as ultimately consisting of loose and separate entities in certain structural relations to each other lacks the means for explaining why the world displays the kind of regularity that we experience. For the central point of this view is to deny that anything more robust *can* be said about the world's holding together and unfolding (rather) regularly over time. Thus Helen Beebe admits that

if there really is nothing in virtue of which the universe is regular, then the fundamental nature of the universe is analogous to the story being played out on the computer screen: it's just a continuous fluke that things go on in the orderly way that they do. (Beebe (2006), p. 527)

From the perspective of common sense, this consequence of a Humean conception of reality seems hard to swallow. Galen Strawson, for instance, argues that part of a realistic outlook of reality is the view that material objects inhabiting this world can affect and modify each other in particular ways, and these events of affecting and modifying are constitutive of what we take causation to be.¹ The common-sense view of causation involves an element of production: If a grain of salt dissolves in water, then water molecules affect its crystalline grid in such a way as to break up the stable structure of the grid. The interactions between water molecules and molecules of the crystalline grid cause the grain to dissolve. Similarly, if a porcelain vase is struck by a hammer, then the force exerted by the hammer affects the fragile structure of the vase, thereby producing its breakage. If this production view of causation is taken seriously, then it apparently contradicts a Humean notion of causation.²

Strawson's critique goes a step farther. A mere production view of causation could also be true in a disordered world where entities act on each other arbitrarily. This, however, does not seem to be our world. Things appear to persist through time and to interact with each other in a regular fashion. The obvious

¹ Strawson (1987). A structurally similar argument can be found in Esfeld (2007).

² Of course, things get even worse once it comes to human action. Within a Humean metaphysics one cannot admit that our experience as agents in the world as being veridical. A similar argument as Strawson's can also be found in Esfeld (2007).

explanation for this observation, according to Strawson, is that things possess and retain certain properties throughout their existence. In virtue of their respective properties, some things can enter into certain forms of interaction with other things. Thus, the assumption that things dispose of a determinate nature is a more probable explanation of the persisting order of our world than the Humean indication that the world's regularity is all just a matter of luck.

Other arguments have recently been advanced which bolster Strawson's argument. Here is one: some philosophers of science point out that the ultimate structure of reality does not preclude dispositional properties. Take electrons, for instance: these don't seem to have an inner structure, but they have spin, negative charge, a magnetic moment, etc. These properties are best described as dispositional; and if electrons indeed are structureless, then these dispositions are not reducible to anything purely categorical. To put it another way, these dispositions are fundamental and ungrounded.³ Although it is reasonable to assume that the dispositions of a complex object — such as the fragility of a vase or the solubility of a grain of salt — are based on the object's microphysical structure, it does not follow that all dispositions are reducible to non-dispositions. It follows only that a rather large number of dispositions of macro-sized objects is reducible to a small number of dispositions at the microphysical level. Molnar summarizes the discussion as follows:

our best credentialled sources of empirical knowledge suggest, to a very high degree of probability, that there are no properties that could serve as putative bases for the powers of the fundamental constituents of the physical world. (Molnar (2006), p. 137)

Interestingly, David Lewis's description of nature's fundamental properties as

perfect natural intrinsic qualities, or of point sized-occupants of points. (Lewis (1999), p. 226)

seems to fit the characterization of an entity's basic dispositions. If electrons lack an inner structure, as current physical theories suggest, then an electron's spin or charge is a (i) fundamental, (ii) natural, and (iii) intrinsic quality which (iv) is instantiated at a point. Nevertheless Lewis did not regard dispositions as belonging to the fundamental level of reality. One major reason for Lewis's rejection of dispositions appears to be his worry that metaphysical speculation might go far beyond what empirical evidence takes for granted; another is his commitment to Hume's denial of any necessary connections in nature. Both reasons may be questioned, however: regarding the first reason, contemporary

³ See Mumford (2007) who speaks of the ungrounded argument.

scientific theory speaks openly about the dispositional nature of ultimate physical particles. Regarding the second, Hume's metaphysical and epistemological reasons for denying any robust notion of causation are highly contestable. There is thus ample room for debate whether we should follow a Humean account of reality.

This brings me to the second argument for Strawson's view: contrary to Hume, we do directly experience causation. I consciously say "causation" instead of "necessary connection" because I think that Hume was right to claim that we are unable to see necessary connections. He was wrong, however, to identify causation with a necessary connection. Briefly put, here is why⁴: In order for a causal process to be truly necessary, cause *A* would have to produce the effect *B* under any circumstances. Causal processes take time, however, and it is therefore always conceivable that a causal process will at some point be interrupted or suspended by an interfering factor so that the effect does not come about. Examples of such cases are legion; nature's causal order is after all untidy.

If this argument is sound, then intermingling causation and necessity results in a distorted view of causation. Anscombe calls attention to this point in her observation that

[e]ffects derive from, arise out, come of, their causes. [...] Now analysis in terms of necessity or universality does not tell us of this derivedness of the effect; rather it forgets about that. For the necessity will be that of laws of nature; through it we shall be able to derive knowledge of the effect from knowledge of the cause, or vice versa, but that does not show us the cause as source of the effect. (Anscombe (1993), pp. 91–92)

In other words, necessity's natural home is the realm of logic; it is a relation holding between propositions. Causation, by contrast, holds between states of affairs in nature. Its modal force is less strong than necessity, for a cause can produce a corresponding effect but its doing so (because of possible interferences) is not necessary. Once the notion of necessity is separated from that of causation, we are in a better position to understand what it might possibly mean to experience causation.

Taking up the view that an entity can in virtue of its causal properties enter a certain spectrum of interactions with other entities, we can conceive of causation as a manifestation of the respective entities' causal properties. An entity disposes of a specific dispositional causal profile which is manifested if the right circumstances obtain. We are aware of this dispositional nature in many cases:

⁴ In Mumford and Anjum (2011), chap. 3., this argument is developed and defended in detail.

we know that we should not light a cigarette at a gas station, for gas has the disposition of being flammable. We know that we should not touch an electric cable, for electricity has the disposition of causing dangerous injuries. We know that a porcelain vase is fragile, so to protect it when shipping it we pack it in styrofoam.

But it is not just the way we think about these and similar situations which indicates that we are familiar with the dispositional causal profile of many material objects. A strong case can be made for the claim that we even have a more direct and immediate access to the world's causal powers. Consider the following cases: You are lost in thought and you bump into a door. You are lifting weights in the gym and you feel how the dumbbells exert a pull on your arms. You experience the effort it takes to cycle up on a mountain. I would like to argue that in these examples we experience the causal powers of various objects firsthand — the hardness of the door, the heaviness of the weights, and the gravity on the slope. Material objects resist our planned activities in various ways, and we feel the effort it takes to overcome this resistance. We even have a specific sensory system for perceiving these causal influences and for responding appropriately to them: proprioception, that is, one's capacity to track one's bodily location, posture and limb position. A proprioceptive system enables us to register the causal influences imposed on us by a material world and to respond to it accordingly. Lifting weights requires a sense of the right amount of effort required: If you exercise too little force, you will not succeed in lifting the weights; if you use too much, you will throw the weights over your shoulder.

To sum up: The Humean claim that, metaphysically speaking, causation is nothing more than a constant conjunction of one entity next to another is highly disputable. I presented three arguments that create room for discussion. First, the Humean view is disputable from the perspective of a realist outlook on reality: it is hard to believe that the rather ordinary course of the world is just a matter of luck and that no metaphysical explanation accounting for this order is available. Second, the Humean view is disputable from the perspective of contemporary science: current theories of particle physics seem to suggest that the nature of the ultimate physical particles is dispositional rather than categorical. Third, it is disputable from the perspective of our experience: an argument can be made that as embodied beings we do indeed experience causation, because the materiality of the world often resists our actions and successful agency requires overcoming this resistance of the material world. From these arguments a further argument, appealing to coherence, may be constructed. Accepting a productive view of causation enables us to connect our commonsense view of reality with scientific findings about the world's physicality and our experience as embodied agents in a material world. Thus, a productive account of causation contributes to the

groundwork for a coherent view of reality ranging from physical particles to our “Lebenswelt”.

2 Events, powers and substances

The previous section contrasted two general metaphysical pictures of reality. It is easy to see that a Humean account is inimical to the very concept of agent causation, for the Humean metaphysics lacks a robust causal understanding. It is thus unsurprising that agent causation is viewed with skepticism by those in the grip of Humean assumptions. In this paragraph I will undertake a closer examination of event causation. I will argue that, once an appropriate metaphysical framework is admitted, a case for reducing event causation to agent causation can be made.

If we consider everyday speech, then we notice that event- and agent-causal formulations are common. We say such things as “Gill caused the vase to break” and “Gill’s hitting the vase caused it to break”, or “The bomb caused great damage” and “The bomb’s going off caused great damage”. The grammatical subject of the verb “to cause” can be an animate human being such as Gill, or an inanimate artefact such as a bomb, or it can be a particular event. Since objects and events belong to different ontological categories, one may infer that ordinary parlance suggests the existence of two different forms of causation — causation by an agent⁵ and causation by an event.

However, standard contemporary causal accounts suggest that agent-causation can be analysed semantically in terms of event-causation, where the former is ontologically reducible to the latter. The idea is as follows: A statement such as “Agent *A* caused event *e*” can be analysed in terms of

I “There is an event *x* which involves agent *A*, and event *x* caused event *e*.” (Lowe (2008), p. 123)

The causal structure of many causative verbs appears to support this analysis. Consider verbs such as “kill”, “put down”, “stop”, “rip”, etc. We may say that these verbs convey the semantic meaning of a means-end-structure. “To kill” means to cause the death of a living being by the use of some means to be further specified; “to put down” means to change an object’s position by ceasing to hold it; “to stop”

⁵ Following the Latin term “agens” I use it in a rather liberal way tantamount to the concept of substance. Thus, if I am talking about “agent causation” it comprises causation by all kinds of substances. There is no need to restrict this concept to animated or even only rational beings.

means to cause the cessation of an object's motion by exercising some force upon it, etc. In these cases a sentence of the structure "A killed B" can be translated event-causally in "A's killing caused B's death". However, this proposal is open to a number of objections.

First, it is widely assumed that certain actions, so-called basic actions, defy such an analysis. Basic actions are not brought about by the performing of some other action but are instead performed directly by the agent. Take, for instance, blinking. If I decide to blink immediately, then I blink immediately. The question by what means I was able to blink does not seem to have any reasonable answer. I may blink in order to do something else, such as giving you a previously agreed sign in a poker game. Blinking can thus be a means toward a further end, but in order to blink I don't have to perform any other action first. Pointing out that my eyelid muscles have to be moved properly for my blinking to take place is of no help here. The proper movements of the relevant body parts is a necessary bodily requirement for blinking, but because these movements are outside our conscious control, they can hardly be described as further actions. On the contrary, it seems more appropriate to say that, by blinking, we cause these movements as physiological realizers of the blinking.

Second, it is a mistake to assume that a noun-phrase referring to an object as the grammatical subject of the verb "to cause" is an elliptic form of speaking because the standard logical form contains a noun-phrase referring to a particular event.⁶ Consider the following sentences:

- "The bomb destroyed the bridge" vs. "The exploding bomb caused the collapsing of the bridge"
- "John lighted a match" vs. "John's lightening the match caused the igniting of the match"
- "Gill opens a window" vs. "Gill's moving her hand caused the opening of the window".

Notice that the sentences on the left are not syntactically incomplete. Translating these sentences into an event-causal form as shown on the right does not imply adding anything that was previously missing. Rather, such a translation requires us to provide additional information about the events involved, for the verb "to cause" does not as such convey the depth of meaning of the verbs in question used in combination with the grammatical subject and object in the left-hand sentences. Generally, we know that a bomb destroys a bridge by causing it to collapse due to the force of the blast, and we are aware that lighting a match

⁶ For a detailed analysis of this and related issues see Keil (2000), pp. 373–383.

amounts to striking it on a rough surface until it ignites. Without this (implicit) knowledge, a desired translation in event-causal terms would fail. The last example complicates matters further. Though we know that a human person has to do something to open a window, it is hard to identify two events here. The movement of the person's hand simply consists in the opening of the window. It is unreasonable to claim that the movement of the hand takes place first and then this causes the opening of the window. A reformulation along the lines of "the moving of the hand happened in such a way that the window opens in the next instant" is at best a clumsy reformulation of the easily comprehensible sentence that a person opens a window with her hand (instead of with the electric window opener), but is never a more lucid explication of that sentence. It is more than disputable that, if someone opens a window, a second event takes place at all. Thus, sentences of the logical form on the left cannot automatically be transformed into sentences of the logical form on the right. The assumption that a sentence conveying causal information can be divided into two events, a prior cause and a subsequent effect, is more than doubtful.

Third, consider the following proposals for translating a sentence such as "By doing *X*, *A* causes *Y*" into "*A*'s *X*-ing caused *Y*'s *F*-ing" or "There is an event *x* which is an action of *A*, and event *x* caused event *y*". Although the proposed translations "*A*'s *X*-ing caused *Y*'s *F*-ing" and "There is an event *x* which is an action of *A*, and event *x* caused event *y*" appear structurally similar, there is a crucial difference between them. If I refer to *A*'s *X*-ing, I remain neutral about *A*'s role in causing an event. It could be that *A*'s *X*-ing happens automatically and so cannot be brought under any intentional action description. If I say instead that event *x* is an action of *A*'s, then *A*'s active role is explicitly confirmed because, for an action to take place, an agent is required who or which brings it about. It appears to be a clear category mistake to say that events can perform actions too. If I am right, then these considerations seem to vindicate the claim that agent-causation ought to be differentiated from event-causation. Consequently, the thing for a proponent of event causation to do is find a way to define the concept of action without any reference to an agent actively bringing it about. "*A*'s action" has to be analysed in event-causal terms with no allusion to the agent at all.

Uwe Meixner discusses two ways how this might be achieved (Meixner (2001), pp. 323–335). Either you aim to spell out a non-causal concept of agency, or you aim to describe the events causing the action as not being causally related to the agent herself. Consider the latter strategy first: if actions are events caused by

other events which cannot be causally attributed to the agent herself, then one must claim that

the causal role assigned to the agent by common sense reduces to, or supervenes on, causal relations among events and states of affairs. (Velleman (2000), p. 130)

In other words, the level of action-involving events is transferred from an agential or personal to a sub-agential or sub-personal level. Causal mechanisms within the agent assume the role traditionally assigned to the agent herself. It is doubtful whether such a strategy is of any real help. A mere transference from the agential to the sub-agential level masks the original *analysandum* without providing any positive account of it. If all we say is that actions are not directly attributable to the agent herself but to action-initiating mechanisms within her, then, still, a positive account must be given of the sense in which this mechanism initiates an action rather than a mere reaction or reflex.

To put it another way: as long as actions are distinct kinds of events, something has to account for this difference. The obvious suggestion within a causal framework is to look at the causal history of actions for identifying the special causal ingredient which turns a causally “ordinary” event into an “action-event”. This suggestion entails a distinction of two different forms of causation: events which cause action-events and events which cause ordinary events. Although we might be in a position to spell out such a distinction in event-causal terms only, a causal dichotomy still lurks which is structurally parallel to the distinction between agent causation and event causation. In either case a special concept of causation is invoked to explain the production of an action.

This brings me to the second strategy. The obvious way to circumvent the challenge of explicating the special causal history of actions is to abandon any form of agent-causal jargon altogether. One might say, for instance, that the causes of an action are within the agent herself but that no special action-initiating mechanisms are involved. To attribute an action to an agent, it suffices to show that the causes are within the agent. Once this approach is pursued, however, it is hard to see how the concept of action can be meaningfully attributed to the *analysandum* at all. As Irving Thalberg remarked many years ago, on such a view the agent mutates into a mere area where

‘his’ calculations, his perceptual judgments, his noble and base inclinations, perhaps his repressed fantasies, his conscious terrors, rages, lusts and devotions, either contend or blend with each other. Even if these proceedings do generate agitation of his limbs, why should we say that this is “his act”. (Thalberg (1980), p. 220)

If the criterion for attributing an action to an agent is only its spatial relation, then we seem to be deprived of any means of identifying the specific difference between an intentionally executed action and a mere bodily motion happening to us. Both events may be “of the agent” in terms of taking place within the agent. But now we are left empty-handed. No proper concept of action is available anymore. The very precondition for differentiating a person’s action from any other form of her bodily motion — that something is up to the agent — is discarded.

These reflections show that the project of reducing agent causation to event causation encounters severe problems on the semantic and ontological levels. If agents perform any actions at all, it is likely that they perform basic actions, and these are not explicable in terms of one event causing another. Rather, basic actions seem to be performed by an agent directly, which amounts to an instance of irreducible agent causation. The aim of reducing agent causation to something non-agential results ultimately in the annihilation of the very notions of agent and action. Agent causation is not reduced to but swallowed by event causation.

In light of these prospects, one might ask whether it might not be worthwhile to try to change the direction of analysis. Instead of explaining agent causation in terms of event causation, one could aim at explaining event causation in terms of agent causation:

II Event *x* caused event *e* if and only if there was some agent *A*, some manner of acting *F*, some agent *B*, and some manner of acting *G*, such that *x* consists in *A*’s *F*-ing and *A* by *F*-ing, caused *e*, which consisted in *B*’s *G*-ing.⁷

This analysis accords well with the outline of the dispositionalist metaphysics drawn in the first paragraph. Consider once more the statement that the exploding bomb causes the bridge to collapse. This statement is true because the following conditions are met: (i) There is a substance, the bomb, and part of its specific dispositional causal profile is to be explosive. (ii) Due to particular circumstances, this disposition was manifested; and its manifestation consists in the bomb’s exploding. (iii) There is another substance, the bridge, disposing of another specific causal profile. This profile includes the disposition to collapse if a strong enough force acts upon it. (iv) The exploding of the bomb is a strong enough force, and therefore the bridge’s disposition to collapse is manifested; this manifestation consists in the event of the bridge collapsing.

This account explains how the two events, the explosion of the bomb and the collapse of the bridge, are causally connected. The connection results from

⁷ This is an expanded proposal of the one suggested in Lowe (2008), p. 136.

substances which, in virtue of their properties, dispose of a causal profile which is manifested under certain conditions. Causation involves the exercise of different substances which act with and against each other in a variety of ways.⁸ It is not hard to see that this account gives an important role to events in causation but at the same time implies that events themselves are causally impotent. Events do not cause other events to happen; rather, events are the result of particulars which interact with each other by manifesting their respective causal dispositions. If a particular's dispositional causal profile were not be manifested, then no event would be instantiated; the reason is that the underlying particular would be inactive. Surprisingly, therefore, this view accords with the Humean claim that events are causally impotent — although its reasons differ significantly from the Humean's. The Humean claim says that causal potency is not a basic feature of reality, for all that there is, are loose and separate events which succeed each other. The account defended here, by contrast, claims that events are causally inert because they are parasitic upon the causal workings of particulars.

Apart from the widespread acceptance of a broadly Humean framework, there is a further explanation for the predominance of event causal accounts which is suggested by Lowe.⁹ He thinks that we tend to resort to this account when we are at least partially ignorant about the real powers at work in a given case of causation. We can claim that event *x* caused event *e* even though we are not in a position to explain in detail why this is so because we are unable to identify the specific substances acting upon each other. Events are epistemically more easily accessible than the underlying powerful substances, and this epistemic accessibility mistakenly involves the inclination to ascribe ontological primacy to events as well as substances. Mumford makes a similar observation when he writes that

we need to distinguish a factive level of what happens in epistemically easy events from a transfactual level of powers that combine to produce those events. (Mumford (2009), p. 108)

The account proposed here holds that agent causation enjoys ontological and conceptual primacy, and it explains how events factor into this picture. Due to lack of space I leave more detailed explications for another occasion. My point is simply that a Humean causal account is by no means the only game in town. On the contrary, there are strong reasons favoring the alternative account of agent causation. I conclude this section by examining three influential objections put forward against it.

⁸ A detailed explication of this view — although in pandispositionalist terms — is found in Mumford (2009).

⁹ Lowe (2008), pp. 138–139.

2.1 The timing objection

The timing objection against agent causation says:

How could an event possibly be determined to happen at a certain date if its total cause contained no factor to which the notion of date has any application? And how can the notion of date have any application to anything that is not an event? (Broad (1952), p. 215)

The kernel of this objection is that reference to a cause should explain why its effect occurred at a given time and not earlier or later. Pointing to a substance *holus bolus* does not provide such an explanation because a substance exists before and presumably also after the explanandum. What appears to be correct about Broad's objection is the claim that an entity's causing something results in that something's happening. A change calls for an answer to the question of what exactly effectuated this change now. However, as we have seen, this observation does not imply that agent causation ought to be substituted by event causation. It is one thing to claim that a particular can only cause by manifesting its dispositions and quite another thing to claim that events themselves are causes. If a substance disposes of a particular causal profile, then an event consists of a substance's causing something due to this profile. Thus, the event depends upon the substance's being causally active in one way or another. Consequently, events are not the right ontological category by which to account for direct causation, because it does not seem proper to say that events dispose of causal powers. This is not to deny that events take place when causation takes place. Rather, it is to deny that events themselves are the causal source bringing about the effect to be explained.

Note that this explication does not claim that a substance as such is the cause of a determinate effect. This claim would indeed be mysterious, because it is hard to see how a substance would manage to bring about an effect without acting in one way or another. Thus, a proponent of agent causation should not say that causation consists in a simple alteration of the event-causal model by replacing the event as cause with the agent as cause. The idea is not that one causal *relatum* is substituted with another.¹⁰ Rather, the claim is that causation is less a relation between two separated entities than a manifestation of a substance's causal profile acting upon another substance (or substances). Causation is not a relation tying together two relata together nomologically, probabilistically or counterfactually; rather, it is the transition of a substance's causal state from

¹⁰ Clarke (2003), p. 186, makes such a suggestion when he writes that “when an agent causes an event, the relation in which the agent stands to that event is the very same one in which one event stands to another when the first causes the second.”

potency to actuality, a transition which entails a range of effects upon other substances. These effects in turn manifest determinate effects depending on their specific causal profile.¹¹

2.2 The objection of non-analysability

The objection of non-analysability claims that the agent's directly causing something remains mysterious because there is no way to analyse the causal relation between the agent and his or her causing. The fact that event causation allows for such an analysis whereas agent causation does not appears to be a crucial objection to the latter. However, we must bear in mind that the concept of the non-analysability is built into the very concept of agent causation. The notion of "directly causing" excludes any internal causal structure and therefore no further causal analysis is available.

Take the example of a radium-atom decaying spontaneously. If we understand that the atom's nature is to decay spontaneously and unpredictably, it is no longer relevant to ask why the nucleus decays now and not at some other moment. Any attempt to divide the event of decaying into two separate further events, one being the cause and the other the consecutive effect, is doomed to fail. All that can be said is that a spontaneous decay is the cause of several effects, such as the emitting of alpha particles and gamma rays. Similarly, one can claim that a spontaneous action is directly caused by an agent itself. Any further causal analysis will lead us down the wrong path, because we will begin to look for items inside the agent to give an action-explanation. This is not to ignore the major differences between an atom decaying suddenly, the neighbour's cat moving spontaneously and my directly bringing about the intention to finish this paper. However, what all these cases have in common is the same basic ontological structure of causation: in each case an entity is endowed with specific causal powers which enable it to produce a range of effects directly, so that no external causal trigger is required in the first place.

2.3 The impossibility objection

One may accept the answers to the two previous objections but advance this one instead: Although agent causation as such seems theoretically intelligible, there is

¹¹ Mumford (2009) defends a pandispositionalist version of this account. He sees causation as a shifting around of different powers as his title of the paper indicates.

a strong reason why it is impossible. You, the proponent of agent causation, claim that substances have certain causal powers in virtue of having certain properties. If a substance's properties were different, then its causal profile would be different too. Consequently, the real cause of the substance's behavior seems to be that substance's having its intrinsic properties, not the substance itself.¹²

This objection works on the basis of a metaphysical distinction between powerful properties possessed by a substance, on the one hand, and the (powerless) substance itself, on the other. Someone may limit the ontological function of a substance to being merely the bearer of powerful properties. Then the substance (or agent) itself vanishes from the causal picture of reality – for only the substances's intrinsic properties relate causally to each other. It is disputable, however, whether the distinction between a substance and its properties amounts to a *distinctio realis*, as opposed to a mere *distinctio rationis*. We are able to draw a conceptual distinction between the substance as mere substratum and the “full-blown” substance with qualities. Yet this distinction does not entail any real, metaphysical distinction. A substance is not a mereological complex entity consisting of simple entities such as a substratum, properties and genuine relations of support between substratum and properties. Rather, a property is a mode of the substance, one of its ways of being. Thus, if we are to drive a wedge between a substance's inherent causal impotence and its having powerful properties, we must assume an ontologically debatable separation between a substance and its properties.

Here my argument that agent causation is a serious alternative to event causation comes to an end. It should have become clear that, once a metaphysical framework of substances disposing of causal powers is established, agent causation fits naturally into it. In the final section I discuss evidence from developmental psychology and cognitive science indicating that the concept of agent causation is not only embedded in a particular metaphysical framework but is also deeply ingrained in our pre-theoretical grasp of ourselves and the world we inhabit.

3 Natural born agents?

As already indicated, the concept of agent causation is generally connected with free and rational actions.¹³ The agent, so an oft-told story goes, has the unique

¹² For this objection see Clarke (2003), pp. 188–193.

¹³ See for instance O'Connor (2001).

power to respond to reasons and to form intentions for actions accordingly. Thus agent causation is essentially intentional and purposive, in contrast to the blind processes of nature which can be reconstructed in event-causal terms. I have argued that there are metaphysical reasons for overcoming this dichotomy, because all real causation consists in substances acting upon each other. In this section I argue that there is epistemic evidence that, from the very beginning, our conceptual system is permeated by the idea of an agent being able to move its body spontaneously. If these epistemic data are true, they correspond with a metaphysics of powerful agents in broad terms and undermine the view to consider only intentional action as agent-caused.

3.1 Agent causation and developmental research

The ability to ascribe mental states such as beliefs, desires and intentions to other people begins rather late in child development and takes years to become fully functional. However, the ability to distinguish between self-moving goal-directed agents and entities in need of an external source of movement emerges much earlier.¹⁴ Research suggests that 6-month-old infants already have a rudimentary capacity to distinguish between humans and inanimate objects in terms of goal-directed movements (Kuhlmeier et al. (2004)).

Spelke, Phillips and Woodward, for instance, discuss a study indicating that infants at this age do not apply what they call the “principle of contact” to the movement of human beings. (Spelke et al. (1995)). This principle says that physical objects move when another object comes into contact with them. In the study, 7-month-old infants were confronted with two different videotaped scenarios, one involving objects and another involving people. In the object-scenario, one inanimate object moved behind a screen and another emerged from the side of the screen. Infants looked longer at this scenario if the second object had begun to move before touching the first object. In the person-scenario, by contrast, a person moved behind the screen and a second person emerged from the side of the screen. If the second person had begun to move before coming into contact with the first, infants showed no signs of more attentively observing this scenario. Rather, they looked longer if the two people made contact first before the second person moved. These findings suggest that 7-month-old infants perceive only people, and not inanimate objects, as being capable of self-propulsion. Other studies complement these findings by showing that 9-month-olds consider the

¹⁴ This paragraph echoes mainly Steward (2009).

self-propulsion of an inanimate object like a robot anomalous, leading to negative reactions and emotional distress (Poulin-Dubois et al. (1996)).

This intriguingly secure grasp of spontaneous human movement in contrast to the motion of inanimate objects is not confined specifically to human agency. Gelman, for example, argues that humans are born with skeletal causal principles which, in combination with perceptual and other cues, lead us to acquire knowledge about animated and inanimated entities in general early on.¹⁵

She calls one principle the “innards principle”. It says that self-propelled agents have insides that enable them to move on their own; she calls another the “external-agent principle”, and this applies to entities that are not in a position to move on their own.

Moreover, infants around the age of one year seem to have a sophisticated non-mentalistic understanding of goal-directed actions which Gergely and Csibra call the “teleological stance”.¹⁶ According to the authors, infants at this age interpret actions as means to an end and evaluate the actions in the light of their efficacy. They can also generate inferences to identify relevant aspects of the action-context which justify the means even if the circumstances are not directly visible to them. The important point for our discussion is that the teleological stance does not involve a conscious ascription of mental states to the agents involved. Rather, it arises from the relationships among three elements: the action, the possible goal and the situational context. Once two of these three elements are given, 12-month-olds are capable of making an inference to the missing element by applying what the authors call the “rationality principle”. This principle assumes that the agent will use the most effective means available in the situational context as the infant perceives it.

Without going into further details, the overall picture suggests that ample evidence supports the view that a basic conception of goal-directed and purposive agency, in contrast to an inanimate object’s mechanistic movement, is part and parcel of our foundational conceptual make-up. One immediate consequence of this distinction appears to be that we directly conceive of animals moving their bodies, presuming that they possess a body which they move in a non-mechanistic way. We do not apply the principle of contact or the external agent principle in order to understand how an animal moves its body. Rather, we apply these principles to inanimate entities: we are inclined to say that there is a

¹⁵ Gelman (1990). See also Setoh et al. (2013) which confirms the assumption that the innards principle goes hand in hand with the ascription of basic biological properties, for instance that even quite diverse animals are not hollow.

¹⁶ Gergely and Csibra (2003).

certain part of this entity, the motor, whose function it is to set the entire object in motion mechanistically. Two further insights accompany this rather “holistic” understanding of an animal¹⁷ and its body. On the one hand, one might say that, once an animal is said to possess a body, then that animal will in some sense be aware of possessing it. The distinction between the animal itself and its body seems to assume that the former has a certain subjective perspective on its body and the world which enables this distinction in the first place. If the animal lacked such a perspective, then one might wonder what supports the claim that the animal is not identical with its body. On the other hand, one might say that, once an animal is said to possess a body, then the animal will also exercise some form of control upon it. And controlling one’s body – even if only minimally – is what grounds the capacity which humans experience as free will. Helen Steward makes this suggestion:

Our natural inclination is to think of an animal as a creature that can, within limits, direct its own activities and which has certain choices about the details of those activities. (Steward (2009), p. 226)

This view converges nicely with the concept of agent-causation which claims, in a nutshell, that the agent has the capacity to bring about her activities directly. Agent-causation, then, not only has a natural home within a general metaphysical framework of powerful substances, but is also ubiquitous among animals. This is not to deny that some entities which we categorize as animals may ultimately, because their movements (contrary to appearances) are reducible to mere stimulus-reaction-mechanisms, turn out not to be true agents. Nor is it to conflate self-reflective rational agency with less complex forms of animal agency. The point is simply that there is no need to restrict agent causation to being the explanans of very special phenomena such as instances of full-blown rational decision-making.

3.2 Agent causation and enactivism

There is further evidence that agency should be considered to be a widespread and basic feature of our existence. Traditionally perception was seen as a passive process, in the sense that sensory input from the world enters the visual system

¹⁷ A traditional substance dualist picture of animated beings is opposed to this understanding of animal movements for it subscribes to a rather mechanistic understanding of how the mind moves the body. The latter is considered as a physical object related to the mind only externally.

and is converted in the brain into a mental image which is, ideally, a correct representation of the perceived object. The perceiver is conceived of, as Alva Noë has put it, as an automatic brain-photoreceptor system whose contents are static snapshot-like retinal images (Noë (2004)). For some time now this conception has been challenged by an alternative picture, so-called enactivism. Simply put, enactivism argues that perception is not a process in the brain whereby the perceptual apparatus constructs a mental representation out of the sensory input provided. Rather, the animal is actively engaged in perceiving, because perceiving itself is a skillful activity performed by the entire animal. This claim is based on the thesis that our perceptual apparatus is essentially connected with our sensorimotor and proprioceptive systems.

To illustrate this point, take vision as a paradigm model of perception. We tend to consider vision as a kind of photographic system: you open your eyes and, thanks to a complex internal process, a focused image of the world in front of you follows immediately. If movement is involved in this model at all, then it is merely as a means of adjusting your perspective in order to gain a better hold of what you wish more sharply to focus on. Moving the camera to the right place and taking the picture are two different events, related only externally. However, there is empirical evidence indicating that this picture is inadequate. Research about blindness, for instance, shows that there are forms of blindness that are not connected with dysfunctions in the visual system as such but rather with the organism's inability to integrate sensory input with patterns of movement. An example is given by attempts to restore vision to patients whom cataracts have made congenitally blind.¹⁸ A cataract is a clouding of the lens of the eye which, in turn, affects vision. If the above camera-model were correct, then removing the cataract would result in removing the thing which impairs the animal's vision. Once the lens is cleared, light passes through to the retina unhindered, which should result in the animal's receiving a sharp image. Interestingly, however, case studies suggest that this does not happen. The surgery restores visual sensation, but this does not automatically restore the ability to see clearly. Immediately after the surgery, some patients continue to suffer a form of blindness. They report that their visual sensations are chaotic, confusing and uninformative to them.

From an enactivist perspective, the plausible explanation is that these patients cannot see because their visual impressions are not coupled with sensorimotor (self-)knowledge. In normal perceivers, sensation goes hand in hand with capacities for movement; we naturally turn our eyes towards an object of interest, we reach towards an object that catches our attention, we reflexively block our

¹⁸ The example is taken from Noë (2004), p. 4.

face with our hands if an object moves towards us. In all of these examples, sensory impressions are automatically coupled with spontaneous movement. One might say that the perceiving subject's visual impressions naturally fit the perceiver's movements because there is an implicit understanding that what is seen depends also on one's own body-posture and movements. The abovementioned patients seem to lack this sort of understanding. They fail to integrate the perceptual objects with their own changing movements (or the ways in which they might move over time), and this failure results in visual impressions which lack any useful content for the perceiver — who experiences this as a form of blindness.

There is further evidence that normal vision itself depends on self-produced movement and concurrent visual feedback. Held and Hein (1963) performed a classical study in which two kittens, one "active" and one "passive" were attached to an apparatus functioning like a carousel with black, white and metal-colored strips on the walls inside. The carousel was moved by the movements of the active kitten who was attached firmly but flexibly to it. The passive kitten was also attached to the apparatus but it was carried in a gondola. It could not move by itself but it was moved in the gondola by the movements of the other kitten. The apparatus was constructed in such a way that the gondola moved in accordance with the movements of the active kitten. The kittens could see neither each other nor their own limbs, but they were able to move their heads freely. Both kittens thus received the same visual input, but only the active kitten, because of its self-movement, received direct sensorimotor stimulation as well. The findings of this experiment are telling: only the active kittens developed normal depth-perception and visually guided paw placement responses. It seems that only through self-movement and concurrent visual feedback can animals develop functioning visually guided behavior. A foundational feature of perception is an implicit practical knowledge of how movements of one's body give rise to changes in sensory stimulation.

If we adopt an enactivist standpoint on vision, then we might not only question the assumption that vision amounts to a passive process of internally representing the world, but we might go a step farther. If an animal is essentially an active embodied being situated in a determinate environment, then why assume that an internal representation intervening between the animal and the world is needed at all? Why not simply suppose that the world is immediately present to the animal? If you want to reach out for a cup of tea, then why assume that doing so requires an internal representation of the cup in front of you? The alternative enactivist account suggests that the very directing of your gaze to the cup amounts to a direct perceiving of the cup as something reachable. The idea is that the cup assumes the role of guiding the hand in your act of reaching for it. In other words, representation may not be required in order for action to

follow on. If animals have fundamentally agential natures, then they may perceive the world directly as full of opportunities for action. This echoes Heidegger's analysis of being-in-the-world which emphasizes that our primary understanding of the world is not one of objects describable in terms of numbers, measures and weights, but of a world loaded with references for use. We perceive the world primarily from the perspective of agents, not observers.

These considerations should suffice to motivate the claim that an enactivist model of perception can support the concept of agent causation proposed in this article. If the agential nature fundamentally shapes the animal's being in the world, then agent-causal terms may provide the most adequate metaphysical reconstruction of this nature. The animal's experience of the world presupposes its active engagement with the world. Any reconstruction in non-agential terms seems to miss the most basic features of what it means for an animal to be alive. This view nicely complements Steward's metaphysical conception of the animal as a self-moving entity which executes some form of direct control over its body. And it might also help us explain why the distinction between animated and inanimated beings figures so centrally in our conceptual scheme. Being self-moving animals ourselves, it is unsurprising that this basic existential feature is mapped into our basic understanding of the world.

4 Conclusion

The contemporary discussion of agent causation focuses on the causal production of free rational action, where such action is seen in radical opposition to the omnipresent event-causal processes which determine natural phenomena. However, if the above arguments are correct, then all causation, whether animate or inanimate, can be modelled along the structural features of agent-causation. First, causation always involves one substance acting on, or being acted on by, another substance. Second, within the animal kingdom agent causation is a mundane phenomenon, because animals themselves are natural-born agents. Third, as a consequence, the production of free rational action is a variation of ordinary animal agent causation brought about by rational animals. One might wonder about the bad philosophical press which agent causation thus far received. Perhaps this reputation has less to do with the concept of agent causation itself than with a long concatenation of philosophical distortions and biases. Substance-dualist worries, empiricist epistemological meticulousness, an overemphasis on mechanistic-reductionist thinking, and a deep mistrust of our commonsense reasoning may have obscured our view of something right under our noses: the fact that, when we experience and interact with the world, we are first and foremost agents.