
Article:
Steward, HC orcid.org/0000-0003-1654-577X (2017) Action as Downward Causation. Royal Institute of Philosophy Supplement, 80. pp. 195-215. ISSN 1358-2461

https://doi.org/10.1017/S1358246117000145

Action as Downward Causation

Abstract: In this paper, I try to argue that the recognition that non-human animals are relevant to the free will problem delivers interesting new ways of thinking about the central metaphysical issues at the heart of that problem. Some such dividends, I suggest, are the following: (i) that the problem of free will can be considered to be just a more specific version of a general question concerning how agency is to be fitted into the natural world; (ii) that action can be usefully regarded as an especially interesting form of downward causation; and that (iii) the metaphysical possibility of downward causation, and hence, indirectly, also of free will, can be illuminated in valuable ways by thinking about the hierarchical structure of, and systems of functioning within, biological organisms.

What is the problem of free will? In this paper, I want to argue that an answer to this question which differs in certain important respects from most of the usual articulations would constitute an important step towards actually solving the problem. The reformulation I envisage, moreover, is no mere change of subject; rather, it is an attempt to show that new resources for tackling even the traditional problem can be helpfully brought into view if one conceives of that traditional problem merely as one facet of a more *general* issue about the nature of agency itself – an issue about what it is for something to *act*, as opposed merely to responding in the manner of an inanimate object to the conditions and circumstances in which it finds itself. I contrast agents with inanimate things because life and agency are importantly related. Perhaps we cannot say absolutely that life is a necessary condition of agency – but at any rate, in my view, the only agents we know of *at present* are certainly living ones. And so, I believe, it makes sense to look to biology, the science which deals with the living, and in particular, to the nature of biological complexity, for help with answers to the question how agency is possible – a question which, I shall argue, is simply a more general (and prior) version of the more usual question how free will is possible. My suggestion will be that if there is to be any hope of providing a metaphysically satisfactory answer to the traditional problem of free will, we must learn to see it as simply a specific version of a broader question about agency and its place in nature. And once we have done so, I argue here, new forms of answer hove into view.

In the first section of the paper, I shall attempt to give a sense of some of the important features of the usual sorts of elucidations of the problem of free will. I have gathered together what I hope is a representative sample of offerings by the simple expedient of sampling the top Google hits which are returned when one searches for the phrase 'The Free Will Problem'. Having drawn out some of the common features of the explanations of the problem which are thus elicited, I shall then move on to suggest, in the second section, that we ought to expand the range of the free will question beyond that assumed by the range of approaches surveyed, so as to encompass all forms of agency and will try to explain why it seems to me indefensible and ultimately incoherent to ask the question only in the limited way in which it has tended to be asked. Then, in the final section of the paper, I shall try to show that thus reconceived, the big metaphysical question at the heart of the free will problem becomes essentially an issue in the philosophy of causation – an issue, specifically, about whether downward causation, understood as the influence of a whole upon its own parts, is possible. I shall try to suggest (though inevitably somewhat speculatively, given the space available) that a proper understanding of the nature of biological organisms gives us reason to think that the answer to this question may be 'yes'.

1. *Traditional Free Will*

According to *The Information Philosopher*, top of the list of Google hits for 'the Free Will problem', '[t]he classic problem of free will is to *reconcile* an element of freedom with the apparent determinism in a world of causes and effects, a world of events in a great causal chain'.[1] Determinism is usually defined as the idea that everything that happens in the world is determined, or settled, by the way things were beforehand, together with the laws of nature, and although I have my doubts about some aspects of this definition, I shall not quarrel with it for the purposes of this paper. I should like instead to focus, for the moment, on another issue – namely, the question of what is meant exactly, in this account of the classic problem, by 'an element of freedom'? What is the element of freedom which the classic problem requires us to reconcile with determinism? Moving a bit further down the entry in the *Information Philosopher*, we find out that this important freedom has to do with 'our will' and 'our actions' – note that interesting word 'our' - and the site then also goes on to mention the moral responsibility we may be supposed sometimes to have for these actions of ours. Compatibilists, we are told, believe that determinism is compatible with moral responsibility – that even if everything is deterministically caused, we can still be morally responsible for at least some of our actions. Incompatibilists, on the other hand, believe that this is not the case, and that moral responsibility depends on the falsity of determinism.

The next Google result on the list is headed 'The Free Will Problem: A Philosopher's Take' and is written by Justin Caouette. Caouette writes that '"Free Will" is a philosophical term of art for a particular sort of capacity of rational agents to choose a course of action from among various alternatives'[2] and that the problem of free will is that this capacity seems to be incompatible both with determinism and with indeterminism – so that it is impossible to have free will, whether determinism is true or not.

The third result is the Wikipedia entry on free will.[3] According to Wikipedia, 'free will is the ability to choose between different possible courses of action. It is closely linked to the concepts of *responsibility, praise, guilt, sin*, and other judgments which apply only to actions that are freely chosen. It is also connected with the concepts of *advice, persuasion, deliberation, and prohibition*'. The problem of free will is then said to arise for those who believe that free will is the capacity for an agent to make choices in which the outcome has not been determined by past events. For determinism suggests that only one course of events is possible, which looks, on the face of it, to be inconsistent with the existence of such free will.

And then the next Google hit - the last one I'm going to consider - is a Youtube clip of a talk by the philosopher Richard Holton, under which the text announces that 'the problem of free will is the question of whether we human beings decide things for ourselves, or are forced to go one way or another'.[4]

In some respects, of course, these four elucidations of what the free will problem actually is are rather different from one another. The second one, for example, suggests that indeterminism may be just as much of a problem for free will as determinism is, something that the other accounts do not mention. The first and third accounts mention responsibility; the second and fourth do not do so. And there are other differences, too. But I am more interested here in something that the four

---

[1] http://www.informationphilosopher.com/freedom/problem/, accessed 19.07.2016.
[2] Justin Caouette, 'Free Will: A Philosopher's Take' at https://aphilosopherstake.com/2012/08/13/the-free-will-problem/, accessed 19.07.2016.
[3] https://en.wikipedia.org/wiki/Free_will, accessed 19.07.2016.
[4] Richard Holton, https://www.youtube.com/watch?v=iSfXdNIolQA, accessed 19.07.2016.

explanations have in common – and that is this. All four of these explanations either say explicitly, or else imply in one way or another, that the free will problem is a problem which specifically concerns *human beings*, and which has no relevance or application to animals other than ourselves. Let us go through them in turn. *The Information Philosopher*, talks of 'our will' and 'our actions' and the need to reconcile these things with determinism. But who is the 'we' that the possessive adjective 'our' is referencing here? I think it is fairly safe to say that it is a widespread convention in the philosophical literature, that when philosophers use the word 'we', they generally mean 'we human beings'. And any doubts we might have had on this score are in any case soon dispelled by the reference to moral responsibility – for moral responsibility seems to be something we can only really sensibly attribute to human beings.[5] And if this is the case, then the *Information Philosopher* entry seems to be suggesting that the problem of free will is essentially a problem about *human* will and *human* action – and the extent to which the element of freedom we generally suppose to be implicit in these things can be reconciled with the doctrine of determinism.

Let us turn to the next of my four examples. Justin Caouette's stated view is that free will is a capacity of rational agents. This is turn raises the interesting question which *are* the rational agents – perhaps if it were allowed that some animals are rational, Caouette's view would not be straightforwardly inconsistent with the idea that animals might have free will. But traditionally, of course, rationality is thought by philosophers to be the distinguishing mark of the *human being* – for Aristotle, for instance, and hundreds of philosophers since, though humans belong to the *genus* animal, our species is distinguished from the others by the differentia of rationality.[6] It seems highly probable, then, that Caouette considers the free will problem, once more, to be a problem about reconciling a specifically *human* capacity with determinism and/or indeterminism. This is, moreover, implicitly confirmed further down the page, where Caouette remarks that free will might also be called 'up-to-usness'. For once again, here we must ask the same question we asked in connection with the entry in *The Information Philosopher*: namely, who is this 'us' whom things are being said to be up to? The natural answer to return is that the implicit 'we' of philosophical discourse is generally the 'we' of humanity – and so we are, I think, justified in supposing that this is what Caouette has in mind, too. Once again, then, the free will problem is being implicitly posed as a problem about a capacity of human beings.

The next case is *Wikipedia*. *Wikipedia* says that free will is the capacity to choose between different possible courses of action. Here, one might think, we have a definition potentially more amenable to a more extended application to non-human animals – for it certainly isn't obvious that the capacity to choose between different possible courses of action is peculiar to human beings. For example, I quite often put two bowls of food down for my cat – one wet food, out of a tin, the other dry food, out of a bag. It does not seem totally implausible to say that when she wanders into the kitchen, she has a choice between at least two possible courses of action – eating the wet food or eating the dry. But whatever impression we might have had that the *Wikipedia* definition of free will is liberal in this regard is immediately dispelled by the next sentence, which tells us that free will is

[5] Of course, we can remonstrate with our pets, and try to train them into behaving as we would wish them to do – but when a puppy chews one's favourite slipper, it isn't really appropriate to think that the puppy is *to blame* in any very deep way. One might punish him, perhaps, to try to stop him doing it in future – but surely no one really thinks that a dog can be *morally* responsible for its actions, even if he can fail to respond as hoped to a training programme.

[6] See, for example, *Nicomachean Ethics*, I 13, which develops the Aristotelian view of humanity as the rational animal species.

closely linked to the concepts of responsibility, praise, guilt and sin – which suggests, once again, a resolutely moral context and a focus exclusively on human beings.

And then finally, there is the Youtube video. At least Richard Holton is absolutely explicit – the problem of free will, he claims, is the problem of whether *we human beings* decide things for ourselves, or are forced to go one way or another. Non-human animals simply do not come into it.

I think, then, that these various websites provide quite a lot of evidence that the traditional free will problem is standardly taken to be a problem exclusively about human beings – a capacity they have which only they have, to choose one course of action over another, and perhaps to do so for reasons, a capacity, moreover, which is very tightly connected to the capacity for moral responsibility. In the rest of this paper, what I want to try to do is to argue that this restriction of the free will problem to human beings has been a serious mistake. Its source, quite probably, is a religious world view, according to which human beings are indeed unique and special creatures, singled out by God for special attention, and given by him a peculiar set of responsibilities, including dominion over the rest of the animal kingdom. We need to try to see, however, whether that view of free will can really survive independently of the support provided by that particular religious context. And I shall be trying to argue that it cannot. I do not wish to deny for a moment that human freedom goes much deeper than the freedoms available to other creatures, because we are the possessors of a range of capacities which enable us to make much more of our freedoms than any other animal can – but we will nevertheless not be able to understand human freedom and its metaphysical requirements properly unless we first think about the animal capacities from which it has evolved.

2. *Free Will as Agency*

One way to see why we need to think about animals in connection with free will is to think about why free will is supposed to be a problem for philosophy in the first place. The traditional issue, as we have just seen from these various websites, is supposed to arise primarily when we contemplate the thesis of determinism, which we are taking for present purposes to be the idea that the future is settled by the past, together with the laws of nature. Now, speaking for myself, it does not seem too too difficult to imagine that a purely *inanimate* world might be deterministic. I don't in fact believe that the inanimate portions of our world *are* entirely deterministic (considered by themselves, and independently of interference by the animate) – for there appear to be a number of truly random – and hence, indeterministic – phenomena in the inanimate parts of nature. Radioactive decay, for example, as currently understood, appears to be an indeterministic phenomenon. Although there are, of course, overall laws governing the phenomenon of radioactive decay, laws which determine the rate at which the overall process must happen for any given radioactive substance, and hence what is the half-life of any given radioactive element, it appears that the emission of individual particles is a random matter, there being no known way to predict or control when such an individual emission event will occur. So in fact, parts of the inanimate world seem to contain indeterministic events. But there does not appear to be anything particularly difficult, conceptually speaking, about imagining an inanimate universe that is deterministic, even if the actual universe is not in fact an example of one. Mostly, we tend to think that inanimate things do what they do simply as a result of the circumstances in which they find themselves, the events which then impinge upon them, and their own intrinsic natures. For example, suppose I add some potassium to water. What happens is that the potassium zips around on the surface of the water and catches alight. Here, we are seeing, no doubt, the operation of certain chemical laws, which govern the alkaline metals, laws which dictate in general respects what will happen when the potassium contacts the water. And

moreover, even though they are doubtless more complicated and difficult to state, most of us would probably think that it was determined not only that the potassium would catch light and zip around, in the manner of the alkaline metals, but also that the precise trajectory of any given, particular piece of potassium was also determined by various prior conditions, in conjunction with more complex and particular laws. Perhaps, for example, the trajectory will be a product of such things as the size and shape of the potassium, the speed at which it hits the surface, the temperature of the water, the shape of the containing vessel, and so on, such that in principle, if we knew all these variables, and how they mattered exactly, we might be able to say where precisely the potassium would go. And this, in turn, is an expression of the conviction that so far as things like bits of potassium are concerned, we more or less expect determinism to be the order of the day, and would be quite surprised if we were to find out that it was not. If the world just consisted of a bit of water and potassium, it seems perfectly conceivable that nothing in the world would be left either to chance or to anything else – that the unfolding of reality through time would be fixed entirely by the properties of those two elements, and the laws which they must obey.

However, the question is, I think, whether this deterministic picture is so readily acceptable once we complicate the nature of the universe we are considering. The traditional philosophical view is, of course, that once human beings enter the picture, there are at least *prima facie* problems about reconciling some of our ordinary beliefs about what human beings are, and the sorts of things they can do, with the thesis of determinism. For we tend to think that on many of the occasions on which we act, more than one course of action is open to us. I could watch TV, or I could go and do some more work. I could walk straight past this homeless person, or bend down to speak to him and try to help. And so on. Whereas if the world is deterministic throughout, it might seem as though these multiple possibilities for action that we think we have are just mirages. The conditions at the beginning of the universe, together with the laws of nature fix or settle exactly what will happen at each subsequent time, rendering free will an illusion, or so the argument goes. Compatibilist philosophers disagree, of course, that determinism would render free will illusory. But for the purposes of this paper, I do not want to go into this debate between compatibilists and incompatibilists. What I want rather to ask is whether the *prima facie* problem of free will must wait for the introduction of human beings into the universe to arise – whether it does not arise already once we add to the world any animals that exceed a certain fairly lowly degree of complexity. Do we really think that absolutely everything that is ever done by a non-human animal is fixed and settled by prior conditions and laws, in just the same way as I suggested we tend to think is the case for things like portions of potassium? Or do we rather think, pre-theoretically, at any rate, that animals are in this respect a bit like humans, with the capacity to choose or decide a certain array of things at the time of action?

I want to suggest that in our everyday thinking, we do not really conceive of many animals in anything like the same way as we think about things like potassium and water. When potassium and water interact, we do not suppose, by and large, that anything is left to be settled at the time of interaction *by the potassium* – it just has the properties it has, and these dictate that it does what it does – and that is that. Whereas so far as the higher animals are concerned, we tend not to think that they do what they do simply as a result of the circumstances in which they find themselves and the relevant laws of nature. In the case of such animals, we tend to posit another factor as well; we are inclined to think that many of the more complex and cognitively sophisticated animals, at any rate, have what one might call a *will*, so that what they do at any given time is partly dependent on *them*, and on decisions or choices they make at the time of action. In talking of decisions and choices, I do not mean to suggest that animals necessarily think things over, weigh alternatives or deliberate, prior to the moment at which they actually act – though there seems in fact to be

evidence that some of them do.[7] In the case of many animals, perhaps action is often undertaken without much, or even any prior thought (as indeed is very often true in our own case). But when something acts in one way though it could have acted in another way, that represents a kind of choice, even if the agent does not think about what to do in advance – a kind of choice I have elsewhere called a *settling*.[8] The animal settles which way the world will go in respect of its body and immediate environs, *as* it acts. But a portion of potassium never settles anything. It just does what the circumstances of its positioning dictate that it will do. The higher animals thus have a distinctive place in our conceptual scheme. They are conceptualised by us as entities which *act*, which are true sources of self-motion, which are possessed of that interesting and distinctive form of spontaneity we call a *will*. They are conceptualised by us, that is to say, as agents. The question now is whether in reality they can live up to the demands placed upon them by this exacting conceptual scheme.

I think a very common worry amongst philosophers is that this view of animals as creatures to whom real options are available is biologically unrealistic. Animals, these philosophers might say, are driven by a set of basic forces which are connected with the need to survive – and that means that they are subject to certain laws of nature, just like everything else. Now, at a certain level of description, and in some limited respects, I am actually perfectly happy to accept this. A hungry and healthy dog, for example, presented with a dish of tempting food, will, in the absence of any particular reason to suppose itself in danger, eat that food. I don't wish to deny that – or any similar obvious truths about what animals of different types will do when confronted with certain exigent circumstances. But there are two very important qualifications. The first is that even if the generic type of activity that an animal will perform in certain sorts of circumstances *is* determined, perhaps by some laws relating to its nature, the *specifics* of its activity need not be. Perhaps it is determined, for example, that the hungry dog will attempt to eat the food. But precisely how fast, how often each mouthful will be chewed, whether this or that portion of food will be eaten first, whether to drink some water in the middle of the whole thing, whether to break off eating to exploit the chance of a walk at some point – these, I suggest, are things which are settled by the dog itself at the very time of its activity – they are not things which have been settled years in advance by circumstances and the laws of nature. Part of the surprisingness, the unpredictability and the chanciness of the world derives from these moment-to-moment settlings by animals of how *precisely* the world will evolve in respect of them – where exactly they will move, and how, at what precise speed and direction, and so on. And the second thing is that not all circumstances *are* exigent. Between feeds, searches for mates, sleeps, and so on, many non-human animals seem to engage, like us, in a kind of leisure – engaging in activities such as sunbathing, grooming, playing, and so on. How long precisely these leisurely interludes between more urgent activity will last, what precisely will go on in detailed respects as they happen, and so on, seems to me most unlikely to have been settled since the dawn of time by the initial conditions and the laws of nature. To think that it is would be to turn what animals do into a kind of clockwork – and their activity into something which is not activity at all, but merely what William James memorably called the 'dull rattling off of a chain that was forged innumerable years ago', the tedious and inevitable unwinding of a set of events to which there is no physically possible alternative.

---

[7] For sceptics, I recommend watching the problem-solving feat managed by a rather remarkable New Caledonian Crow at https://www.youtube.com/watch?v=cbSu2PXOTOc. It is almost impossible not to ascribe a deliberative thought process to the crow when watching it perform this task.

[8] See my *A Metaphysics for Freedom* (Oxford: Oxford University Press, 2012) for the development of this concept.

This brings me to a point which I think is of considerable importance in the philosophy of action – and that is that in my view, and for the sorts of reasons I have just given, it is the phenomenon of action *itself* which is in ostensible tension with deterministic visions of the world. It is important to emphasize how different this is from the view which appears now to be standard in the literature on free will. The standard view is that action itself is a widespread phenomenon which is deterministic in many of its manifestations, and in particular is deterministic in all its *non-human* manifestations – but there is a special kind of action – usually referred to as *free* action – which is such that the agent, though she in fact does some particular thing at a given time, could have done something else instead. In other words, the metaphysical picture of the world of actions is something like that pictured in Figure 1:

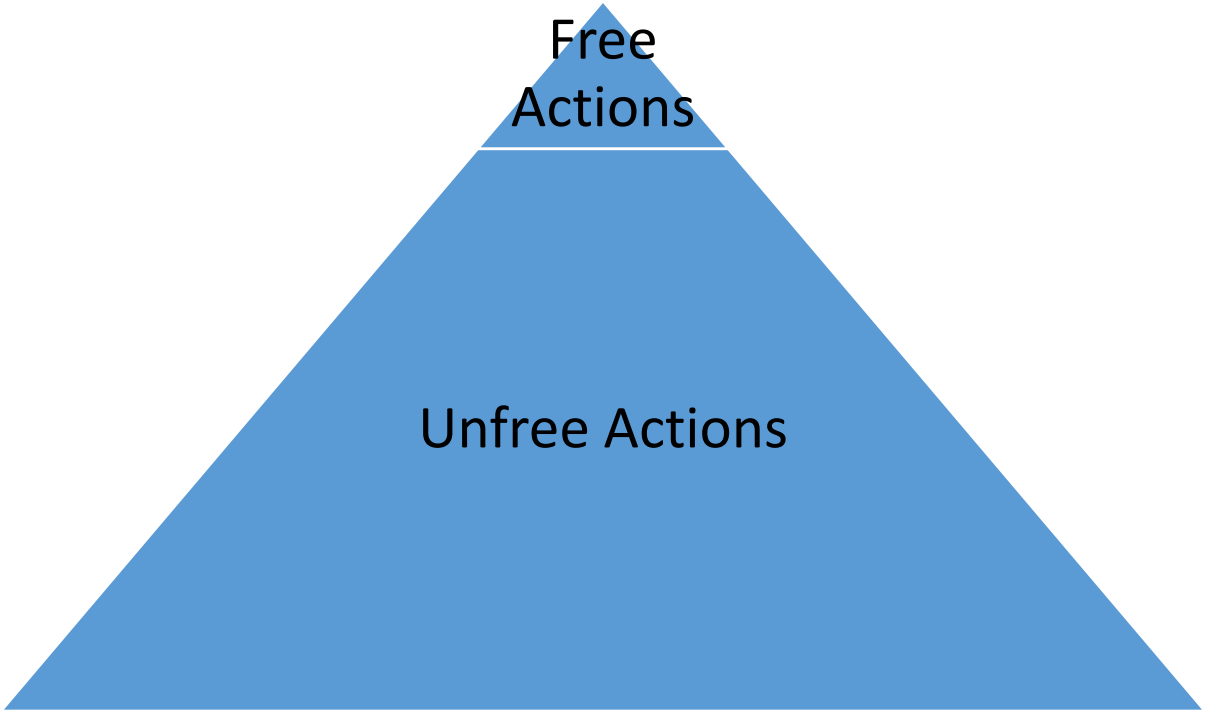

Figure 1: Actions and 'free' actions on the traditional view.

These free actions, according to the standard view, are only ever performed by human beings – all other actions (including some human ones) being unfree. But in my view, action is a phenomenon which is *always* indeterministic in its manifestations, because part of what it is for an agent to *act* is to exercise a power at a given time when she needn't have exercised it then. This does not mean that she cannot also be exercising powers which she *has* to exercise then, as a matter of some sort of law, or imperative of nature, perhaps a biological or evolutionary one. Think back to my cat, with her dry food and her wet food. Perhaps, if she is hungry enough, and nothing is putting her off, there is some sense in which she's simply *bound* to eat the food once she notices its presence. But is she also bound to eat the dry food first? Or, if she is, is she bound to chew it by making *precisely* the motions she does in fact make? Is she really determined to go through *precisely* the set of motions she does in fact go through in eating the food, and has the fact that she will go through precisely this set of motions been settled since the dawn of time? My answer to this question is that this is a very unnatural thing to think, and moreover that it seems inconsistent with thinking of the cat as a proper

agent. This picture transforms the cat from a being with agency into a mere machine, essentially a mere *place* in which various inevitable interactions occur. Part of what is involved in the cat's activity actually being a true action of the cat's in the first place is the thought that the cat didn't have to do exactly what it did. An action, on this view of agency, *just is* the bringing about of some movement or change in the universe *by an agent* – and I think it is very hard to see how an agent could ever be the true *source* of any such movement or change if happenings within her are the entirely deterministic causes of those movements or changes, and if those happenings within her are themselves merely the inevitable consequences of the past and the laws.

I think philosophers of action have often had trouble accepting this indeterministic view of agency, because they are operating from the start with an incorrect view of what actions *are*. A very standard view – often called the Causal Theory of Action – holds that actions are just bodily movements caused by certain sorts of mental states – things like beliefs and desires, and intentions. And if this is your view of action then it is unsurprising that there does not seem to be any problem about reconciling the existence of actions with determinism – for on this view, the existence of actions actually *requires* that tight causal relations exist between prior mental states, on the one hand, and bodily movements on the other. Of course, one might wonder whether the causation involved is always deterministic – many people accept these days that there can be merely probabilistic forms of causation – but even if we admit that, it is hard to see how an indeterministic nexus between mental states, on the one hand, and bodily movements, on the other, could really help us understand agency. As compatibilist philosophers never tire of pointing out, if my desires and beliefs lead causally to my action but only with a certain high degree of probability, that seems to make things *worse* for freedom, not better; as Laura Ekstrom once put it, that would be a view on which I seem to have to 'wait to see' whether I will act as a result of an intention to do so, and that surely cannot be the right way to ensure that the agent truly gets in on the act when an action occurs.[9] But my worry is that on a standard Causal Theory of Action, it is not clear how an agent can *ever* really get in on the act. The things which seem to be truly causally efficacious, on the Causal Theory of Action are states and events – not agents. On the Causal Theory of Action, when an agent acts, it is only by virtue, as it were, of states and events usually conceived of as being inside her – perhaps inside her head –causally interacting in various ways. But my inclination is to think that this picture of action rather loses the agent altogether, turning her into a mere *place,* a location where various events occur. My desires and intentions aren't *me* – they are merely properties of me – and so *their* causing things needn't be the same thing as *me* causing things. The view of action that I favour offers an alternative to this Causal Theory. On my view, actions are intrinsically linked to agents – they just *are* events (or better still, processes) which are the settlings by agents of a range of questions to which the answers have not yet been settled – questions such as 'where am I going to be at time t_1'? 'at what speed will I be moving at time t_2'? 'Will I be eating or not at time t_3?' etc. (though of course I don't mean to suggest that the agent ever has to explicitly consider any of these questions herself in order to count as having settled them). But if actions are to be such settlings, it is essential that more than one possibility exist for the agent at the relevant moment – how otherwise could it possibly be the case that the agent settles anything *at* that moment? One cannot settle at time *t*, what has already been settled in advance of time *t*. So agency requires indeterminism necessarily. There is, though, nothing intrinsically special to *humans* about agency as I have characterised it here. If an action is just the settling by an agent at the time of action of the answers to a range of questions to which the answers are not yet settled, then there is no reason to think it is a capacity restricted to human beings. Though there are interesting questions about how

[9] Laura Ekstrom, *Free Will: A Philosophical Study* (Boulder, CO: Westview Press): 105.

far down the scale of complexity the phenemenon of agency may be supposed to extend, it seems evident that dogs, horses, dolphins, and many other animals are certainly agents. And that implies, on the view of action I want to embrace, that they having the capacity to settle through the process of their actions, and at the time of those actions, that the world will go one way, when it could have gone another.

There are of course many possible objections to the view I have just tried to outline – and it will not be possible for me to deal with all of them here.[10] What I want to do in the remainder of the paper is just to look at one of them, and to try to explain how this new framing of the free will problem might help to deliver the outlines of a solution to it.

3. *Action as Downward Causation*

Consider the phenomenon of bodily action. In acting, I make my body move in certain ways – for example, I raise my arm, I bend my leg, etc. – and perhaps by means of these bodily movements, I bring about further effects in the world. But my body cannot move in these various ways unless certain things first happen in my brain and central nervous system – for example, certain neurons must fire in my motor cortex. It would seem, then, that in order to bring about the resultant bodily movement I must either bring about the prior activity in my motor cortex as well – or the activity in my motor cortex must simply *be* (at least part of) the process which constitutes my bringing about the bodily movement in question. If we choose the former answer – that I must bring about the activity in my motor cortex - the question merely arises again how is it possible for me to bring about *this* activity – and the answer would seem, again, to be the disjunctive one that it is possible only if I am able to bring about still *prior* neural activity which produces the activity in the motor cortex; or if the prior neural activity is at least part of a process which *constitutes* my bringing about the activity in the motor cortex. And if we choose the former answer, once again the question will be raised: how on earth is it possible for me to bring about this prior neural activity – and again, the same disjunctive reply seems inevitable. We do not seem to be able to end the impending regress without at some stage either concluding that the whole chain of neural activity must be initiated ultimately by an ethereal input from something like an immaterial self which sets off a whole chain of physical causes, as in Figure 2; or else that my activity is at the end of the day entirely constituted by the activity of certain of my functionally significant smaller parts on other such parts – neuron on neuron, synapse on synapse, and so on.

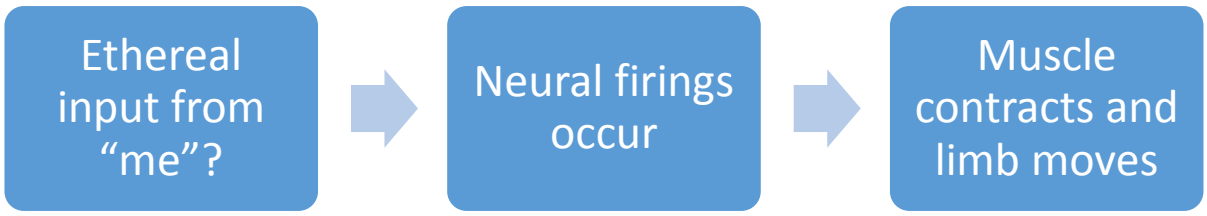

Figure 2: Ending the regress: the dualist's picture

[10] I have dealt in some detail with what seem to me to be the most considerable of them in my *A Metaphysics for Freedom.*

Since along with many other philosophers, I take the former dualist solution to be unacceptable for a variety of reasons, *some* version of the latter must be the right thing to say. In some sense or other, my activity must always be realised in the activities of parts of my body – there is no acting on my part, which is not realised in some way by these lower-level events. Action, after all, is not magic – it needs a physical realisation if it is to create physical effects such as bodily movements. But the question is whether it is possible to say this and yet avoid the conclusion that it is not really me but rather my *parts,* and the events that are occurring in them, that are doing all the important causal work. If my actions are simply constituted by neural activity, where am I to be found in the causal story? It is hard to see how, if the story is correct, the agent herself can be anything more than a kind of epiphenomenon, arising out of the hive of activity taking place in the cells, muscles, blood vessels, etc.. And it is hard, also, to see how determinism can be avoided. For don't the activities of nerves and muscles have local and deterministic causes? And if they do, how can the actions which result from these activities of nerves and muscles possibly avoid capture in the deterministic web of events?

What seems absolutely essential, if we are to avoid the spectre of a dualistic input at the beginning of the causal chains which in one way or another underlie our actions, is that we avoid thinking of an animal's input into the course of nature as something *prior* to whatever neural processes initiate and then monitor and control the relevant bodily movement or change in the causation of which an action consists. For that just leads to the dilemma already discussed - either the prior input by the animal is itself a neural process, in which case we just face the same question again about how that prior neural process has been produced by the animal – or it is not, in which case it is hard to see how dualism is to be avoided. The key, if we are going to make proper room for animal agency must be to see the animal's input as a matter not of *prior* intervention but of *top-down* control, control of at least some of the processes taking place within and around certain small physical things (such as neurons and synapses) by a larger one, the animal which those small things partly comprise. What it seems we have to be able to make intelligible to ourselves is the possibility that a whole animal might have top-down effects on its own parts.

But how can a whole thing affect its own parts? Many of us are so used to thinking reductively about complex entities, that this might seem, at first, to be simply impossible. We are used to thinking that the macroscopic behaviour of complex things is ultimately due entirely to the activities of the small bits and pieces of which they are made. This, for example, is how we tend to think about a washing machine, or a TV set. What happens with the drum of the washing machine, or the screen of the TV is, we think, due to events going on inside it of which most of us have only a fairly dim understanding. The behaviour of the whole entity, we tend to think, is dictated by the behaviour of its parts, the whole being an immensely complex mechanism in which overall outputs, given any particular input, are determined by a certain arrangement and disposition of internal circuitry. But we should not allow the importance of this sort of bottom-up determination of large-scale effects by small-scale transactions to blind us to the fact that influence may also flow in the opposite direction. The key to the understanding of how this is possible, I shall argue, lies in the two phenomena of *coincidence* and *ordering*.[11]

In general, a great deal of what happens in our universe is able to do so only because of various forms of spatial and temporal *ordering.* When molecules are connected together in certain spatial arrangements, to form a macroscopic physical object with a distinctive set of powers, it provides us with what is perhaps the simplest example of this sort of phenomenon. Roger Sperry, a neuropsychologist, was a passionate defender of the idea of downward causation which he regarded as essential in order to account for consciousness. In Sperry's view, though,

[11] See also my *A Metaphysics for Freedom*, Chapter 8, from which the main lineaments of some aspects the argument which follows is taken.

downward causation was not just a phenomenon peculiar to the mental realm – but on the contrary, a quite ubiquitous occurrence. The example of downward causation he often uses is that of a wheel. When a wheel rolls downhill, he notes, 'the molecules and atoms … are carried along … regardless of whether the individual molecules and atoms happen to like it or not'.[12] Sperry's point is that although the wheel is composed of molecules and atoms, whose particular features doubtless determine certain macroscopic features of the wheel (for example, how flexible it is, how strong, and so on) it is also true that certain macroscopic features of the wheel (in particular, its *shape*) determine what will happen to the individual molecules of which it is composed, given that the wheel is placed in certain circumstances (e.g. on an inclined plane). The individual molecules and atoms in the wheel can only move in ways which are enormously constrained by their being bound up into a particular kind of larger whole.

It might be wondered whether this really counts as downward causation. Jaegwon Kim, for example, has argued very forcefully that the idea of downward causation does not, in the end, make sense, because the effects one might be inclined initially to attribute to macroscopic or higher-level phenomena seem on reflection to be re-assignable to the microscopic or lower-level ones which together give rise to the higher level state-of-affairs in the first place.[13] As Kim puts it: 'the difficulties [with downward causation] essentially boil down to the following single argument. If an emergent, M, emerges from basal conditions C, why can't C displace M as a cause of any putative effect of M? Why doesn't C do all the work in bringing about the putative effect of M and suffice as an explanation of why the effect occurred?'[14] Kim's question put in terms of Sperry's particular example, is how on earth the *wheel* can have any causal role to play, over and above the causal role played by the molecules of which it is composed. How, one might ask, can the wheel possibly be an extra causal factor, instead of just being displaced, as a causal player, by the molecules which go to make it up? Those molecules and the bonds between them, one might say, produce the wheel shape in the first place, and so any effects attributable to that shape are actually attributable, in the end, to the molecules arranged wheel-wise. There is therefore no irreducibly downward causation here – all causality flows down inexorably to the lowest level, of which all the rest are revealed merely to be the upwardly determined upshots.

In my view, however, this argument is too quick. Is it really true that the molecules and the bonds between them 'produce the wheel shape in the first place'? It is true that *once the molecules come to be arranged as a wheel*, then of course, wheel-specific effects will follow, which are supervenient of the lower-level arrangements. But of course molecules do not tend spontaneously to form themselves into wheels. Wheels have to be *produced* – and for that, one requires a great many sorts of coincidences and orderings to occur. One requires, for example, a wheelwright with the requisite skills, intentions, tools and raw materials. These things have to come together in the right place at the right time, and then the wheelwright must then use his skills to act upon those raw materials so as to produce a wheel – a process which will in turn require that an enormous number of brain processes take place in the right way and in the right order inside him – processes relating to visual and tactile perception, to motor skills, to memories relating to previous efforts to make similar items, judgements about how to solve difficulties relating to the idiosyncrasies of these particular materials, and so on. In other words, in order for molecules to *become* arranged in such a way that they come to constitute a wheel, an enormous number of separate events must occur together and/or in the right sequential order. But how does the world provide for this coincidence? What is the causal story of its production? The need for such a story puts pressure on the idea that the relevant causation is to be understood wholly as a matter of various forces blindly acting on such things as molecules and atoms, and may

[12] Roger Sperry, 'A modified concept of consciousness', *Psychological Review* **76** (1969): 532-6.
[13] Jaegwon Kim, 'Making sense of downward causation', in P. Andersen, C. Emmeche, N.O. Finnemann and P. V. Christiansen, *Downward Causation* (Aarhus: Aarhus University Press, 2000): 305-21.
[14] Kim, 'Making sense of downward causation', 318.

thereby simultaneously help to provide us with the answer to Kim's question. For part of the causal story about this particular set of molecules and their journey through the world will involve the telling of the causal story about the coming-into-existence of a wheel – and the causal explanation of *that*, one might argue, cannot be given entirely in terms of lower-level phenomena. In particular, a great many factors are required to *coincide* and thence to *form orderly sequences*, if a wheel is ever to come into being – and it seems very difficult to understand how the requisite coincidences and sequencings have been brought about if we stick doggedly to describing phenomena at the molecular level of resolution. How on earth has it come to be that all the various molecular phenomena which need to coincide in particular, distinctive ways in order to constitute e.g. the existence of a wheelwright with a certain intention, and the existence of various tools, and so on, have so fortuitously arranged themselves! For the answer to this question, I suggest, we need to raise our eyes from the atomic and molecular and look to the realm of the macroscopic – for answers which in this case involve the existence of persons with plans and ideas about how to get them enacted. In Kim's terms, then, M (the wheel) thus displaces C (the basal conditions), because C would never have come about in the first place were it not for the fact that C constitutes M, given that M is the thing that is wanted by an intentional agent.

One might object to this line of thought that the requirement of manufacture is merely a contingent and unnecessary feature of the particular example used by Sperry. If Sperry's example works, someone might suggest, it should work for, e.g., a rounded boulder, just as well as for a wheel – a case in which the sorts of complications rehearsed above, which are to do with intentional creation, would be absent. A boulder, presumably, acquires the eventual shape it does because of a range of historic interactions with other objects and stuffs – things such as glaciers, other rocks, water, and so on. But even in this simple case, the idea that a decent causal account of these interactions can be given without appealing to forces which operate on the *macroscopic* objects in question, in virtue of their strictly macroscopic features seems questionable. A boulder which begins life in the sea when it falls from a sea-cliff might, for example, grow gradually rounder over the course of many years because of the way in which it is dashed against the cliffs by the waves. But how precisely it will wear as a result of this constant dashing seems to be (in part, anyhow) a matter of such macroscopic matters as its original shape, the macroscopic shape of the objects it is dashed against, and so on. It is true, of course, that these shapes 'supervene' on various arrangements of molecular entities. But it is not clear that a *causal* understanding of how the molecular arrangement which constitutes the boulder came to be in the first place can be properly provided entirely at the molecular level. In an admittedly rather 'thin' sense, it appears, even in this simpler case, as though macroscopic features of *the boulder* might matter to the causation of effects on its own parts. The way in which its parts are *ordered* and the way in which the dynamic interactions which produce its eventual shape come about is a story essentially about macroscopic forces, not microscopic ones.

In the case of the boulder, then, I think we could say something like this: that there is a sense in which a very limited sort of top-down causation exists in virtue of the relevance of features of the boulder taken as a whole object to subsequent changes undergone by the boulder, changes which in turn affect what is true of the boulder at the *molecular* level. None of these changes, of course, can take place independently of changes which themselves have molecular descriptions; and there is certainly nothing here to disturb the hypothesis of determinism. The point is merely that the laws and principles in terms of which the changes are to be understood are macro-laws, not micro-ones. No doubt if this is downward causation, it is downward causation of a rather limited sort. But it is a starting point from which we can perhaps begin to see how an understanding of the nature of a whole may be requisite for a full causal account of the trajectory taken through the world by the parts of that whole. It is when we enter

the realm of the biological, and in particular, of the psychological, that the phenomenon of top-down causation really comes into its own.

Where animate entities are concerned, the importance of these phenomena of coincidence and ordering becomes much greater than in the case of merely inanimate objects – and the dominance of whole over part is, relatedly, much more significant. It is arguable, indeed, that a certain hierarchical holism is quite ubiquitous in biology. Any animal larger than a single cell is a hierarchically organised entity – cells are organised into tissues, tissues into organs, organs into systems (e.g. the digestive system, the circulatory system, the visual system), and all these systems are organised in their turn so as to operate together for the benefit of the whole organism of which they are the subsystems. And there are ways in which, at every level of this biological hierarchy, entities at the higher level dominate and constrain processes occurring in the lower level ones. Even a single cell is a structure, for example, which, once formed, exercises a certain sort of dominance over the processes which go on inside it.[15] None of the individual processes which constitute the life of a cell is independent of the others – and it is to the functional needs of the tissue, organ, and ultimately the animal, of which the cell is a component that we must look for an explanation of why the particular processes which co-exist together in the cell have thus come to coincide there, and why they take the specific forms they do. The organisation of a cell thus cannot be understood without considering it as embedded in the functional units of a whole organism.[16]

I believe we should conceive of animal agency as an essential part of this hierarchy of domination-relations – as the power which belongs to certain sorts of whole organisms to organise the operations of certain of the various sub-systems at their disposal, in such a way as to benefit them, as they confront the contingencies of life. Where action is concerned, the requirements for various coincidences and orderings to obtain are vast, and not all of them can be planned for in advance by the instigation of mere instincts and habits. Take, for example, someone's swimming a few lengths of a swimming pool. For a start, the swimmer's arms and legs must be co-ordinated – the movements or each limb must occur at the right time if the stroke is to propel the swimmer forward effectively. Then the swimmer's breathing has to be controlled in such a way that in-breaths occur when the swimmer's head is above the water and out-breaths when it is below. Learning to produce these forms of control and co-ordination habitually is a crucial part of learning to swim for human beings; whereas for other animals, it is part of an instinctive endowment. But not everything needed for swimming can be handed over to these sorts of habitual or instinctive mechanisms. The swimmer must, for example, prepare to turn as she sees the end of the pool approach. She may need to take account of others in the lane: to speed up to overtake someone, or to slow down, to permit someone else to overtake. If she sees a friend in the pool, she may need to break off her swim to say hello, to avoid giving offence. If the fire alarm goes off, she will need to be able to understand what it signifies and abandon the swim altogether. And so on. Sub-systems, in other words, need to be co-ordinated on the spot by an overall co-ordinator which is able to respond to the unexpected, the unpredictable, the contingent, the accidental. A*ction* thus emerges when the need for *discretion* enters the biological hierarchy – when a creature itself evolves the power selectively to control certain of its own sub-systems in the light of incoming information, in such a way (roughly) as to optimise its chances of survival and success. In particular, higher animals need to have this kind of discretionary control over their own locomotion – since it is often decisions about where and

[15] See Donald Campbell, '"Downward causation" in hierarchically organised biological systems', in Ayala and Dobzhansky (eds.) *Studies in the Philosophy of Biology* (Berkeley and Los Angeles: University of California Press, 179-86).

[16] For detailed accounts of some of these cellular processes, see A. Moreno and J.Umerez, 'Downward Causation at the Core of Living Organisation', in P. Andersen, C. Emmeche, N.O. Finnemann and P. V. Christiansen, *Downward Causation* (Aarhus: Aarhus University Press, 2000): 99-117.

how to move upon which survival depends – decisions about what to chase, what to flee from, where to hide, and so on. For a mobile creature with many needs, and many competing ends, some way of integrating the operation of these various systems so that the right range of things can be done in a sensible order, must be instigated; nature has found that habits, instincts and tropisms will not always suffice for the survival of a complex and self-moving creature. What it has instead found, I surmise, is that the type of system which best serves those needs is precisely the type that I have here called an agent – a creature that is a settler of matters concerning certain of the movements of its own body, and on whose discretionary settlings its own persistence and flourishing depend.

I have attempted, then, to argue in this paper that action should be thought of as a special form of downward causation. This view has the enormous benefit of placing action into a broadly naturalistic, biological context, which can be seen as in some ways continuous with the other forms of downward causation which are found in hierarchically-organised systems. But the view also gives due recognition to the extreme specialness of action, in that it recognises the discretionary as a genuinely new and emergent phenomenon of life. In that sense, it accepts and respects the presupposition of the traditional free will debate that free will is a metaphysical conundrum – something distinctive in the order of nature which requires a special explanation. Were I to speculate, I should propose that some of philosophy of mind's *other* conundrums – in particular, consciousness and self-hood - also come into being, evolutionarily speaking, alongside the development of discretionary agency. There is thus a prospect, in this biological perspective, I believe, of uniting aspects of the philosophy of mind that tend to be treated entirely in separation, but ought not to be.

Helen Steward

University of Leeds

h.steward@leeds.ac.uk