# 2
# Agency

## with Helen Steward

*Helen Steward*: I first became interested in the power I call 'agency' through thinking about the problem that determinism is supposed to present for free will.[1] Not everyone agrees on how exactly free will should be characterised. For some it involves the power to do otherwise than one in fact does; for others it attaches to loftier notions like self-determination and autonomy.[2] But there seemed, at the time, to be general agreement that free will should be thought of as a power specific (at any rate so far as its manifestations on earth are concerned) to human beings – and that the 'problem' which determinism presents is one which arises only in connection with certain human-specific powers. In particular, virtually none of the literature on free will ever mentioned the capacities of non-human animals.

But to me, the idea that a universe containing the activities of creatures such as chimpanzees, dolphins, elephants and dogs might safely be regarded as entirely deterministic seemed as strange as the idea that one containing human beings might be so. Surely, I thought, at least some animals do things in contexts in which it is right to say that they could have done otherwise (than do those things). Think of a cow meandering around a field, going this way and that; or two squirrels rolling and tumbling around in play; or birds selecting nesting material. In all these cases, and many others like them, it seemed to me, we perceive the exercise of various powers which seem, on the face of it, just as difficult to square with determinism as do the morally loaded human choices on which the philosophical free will literature tends to focus. I felt convinced that a certain degree of freedom attached to all these animal activities – and I was interested in exploring further whether that thought could be justified.

It is not surprising, of course, that the free will debate has tended, historically, to focus on capacities unique to human beings. That debate, like most other longstanding philosophical debates, has its roots in a religious context which took the special status of human beings entirely for granted – they were creatures made in God's image and so creatures which of course should be expected to have unique metaphysical powers, and to be 'outside' nature in certain respects. But stripped of that religious context, the supposition that it is human agency *alone* that presents problems for determinism looks much more problematic. What on earth could it be about us humans, specifically, that presents an obstacle to determinism? What is it about our mode of functioning *in particular* that makes it look as though a certain kind of flexibility must be present in nature? Libertarianism can come to look foolishly anthropocentric and anti-naturalistic when it alleges that it is *just us*, alone in the natural world, who have managed to escape the shackles of an otherwise thoroughgoing determinism. I was interested in the prospect of developing a more naturalistic style of incompatibilism that might attempt to root a determinism-disrupting freedom somehow in animal biology. We already have reason to suspect that life introduces surprising new properties to the world – I wanted to argue that agency might be one of them.

A concept I have made use of in the development of my ideas about agency is the concept of 'settling'.[3] The vision of the universe that determinism offers us is one in which everything that happens is settled in advance by things that have already happened or which are otherwise already fixed (such as the laws of nature or God's prescription). If determinism is true, nothing that happens is ever newly settled at the time at which it occurs – since everything that happens is *already* settled – and nothing can be (newly) settled *at t* that was already settled prior to *t*. But I think our natural conception of action supposes actions themselves to be settlings of the hitherto unsettled – things such that it is up to the agent *at the time of action* whether or not they will occur. Actions simply cannot be settlings in this sense, if determinism is true. And so if I am right that it is in fact essential to the concept of an action that it be a settling by its agent of what is hitherto unsettled, then there is a conflict not only between human-specific powers of 'free will' or 'freely willed action' and determinism, but between agency itself and determinism. Actions themselves then cannot be accommodated by the deterministic universe. This view I call 'agency incompatibilism' and I think it offers a much more palatable version of libertarianism than do many traditional versions of incompatibilism.

*Is* it right to think that it is essential to the concept of an action that it be a settling by its agent of what is hitherto unsettled? There are of course conceptions of actions on which this wouldn't be true. Some think, for example, that inanimate entities can act – that water acts, for instance when it dissolves salt.[4] I don't mind, of course, if people want to use the language in that way. But I do want to insist that there is another concept for which no better term suggests itself than 'action', which is such that water is not the sort of thing that could be engaged in it. Nothing is ever up to water. When water 'acts' on salt, what is occurring is an inevitable interaction, and there is nothing at all strange about the supposition that all of the fine details of this interaction are settled in advance of its actual occurrence (though as a matter of fact, the supposition could be false),[5] by facts about the distribution of the salt, the temperature of the water, the shape of the container in which the water is held, etc. My interest is in a conception of action and agency which draws a fundamental distinction between the 'doings' of things like water and another class of doings, the ones which manifest what I should like to call the *two-way power* of agency.[6]

*

*Michael Hauskeller*: It is indeed important to distinguish between the way an inanimate object can be said to 'act' or do things and the way we humans and also many (if not all) animals do. Ordinary language tends to blur or even ignore that difference.[7] I can be hit by a rock, but I can also be hit by another person, but what the rock does is in fact quite different from what the person does when they perform an act of hitting. It is an altogether different category, a different *sense* of doing, and what makes it different clearly has something to do with the difference between the power we think rocks have (and the power we think they lack) and the power both people and many non-human animals appear to have. 'Two-way power' is a good term for this: it is the power to do *or* not do a certain thing. In contrast, inanimate objects like rocks have only one-way power: the power to do a certain thing without the power not to do it. When prompted in a certain way (e.g. you pick it up and throw it), the rock will act in a certain way (i.e. follow the trajectory of your throw and hit me). It has the power to do so, but not the power not to do so. We, on the other hand, and generally all proper agents, also have the power when prompted in a certain way (e.g. I annoy you) to act in a particular way (e.g. you throw a rock to hit me) or not to act in that particular way (i.e. not throw it and do something else instead). In other words, we have the power to act in *more than one* way. Or so it appears to us.[8]

Yet even though we strongly feel that we have the power to do otherwise than we in fact do, we are also very much aware that this power is far from unlimited. What freedom we have, we have within the limits of our nature. We can be fairly sure that there are certain actions that somebody we know (be it a human or a non-human animal) is not going to perform. It is extremely unlikely that the colleague I'm having a relaxed chat with during my lunch break will suddenly start screaming and destroying the furniture, or that my beloved dog will suddenly, in the middle of her morning walk, turn on me and tear me apart. Clearly, this is not physically impossible for them and it wouldn't break any known laws of nature if they did it, but, *being who and what they are*,[9] it is just not something they would do, and if they did do it we would immediately assume that they must have lost control over their actions and that something else must have stripped them of their agency and made them behave in that otherwise inexplicable manner.[10]

In that sense their actions are more or less predictable.[11] They are free to do certain things or not to do them, but they are not really free to do just about anything or not to do it. There are options, but those options are quite limited and determined by the situation and the nature and history of the agent. This is why there is a certain degree of reliability in our interactions with other agents. Accordingly, agency is in practice not the power to do something that goes against our nature. It is the power to do a certain range of things that lie *within* our nature to do (and to not do the things that do not lie in our nature to do).

But how do we know that we have such a power? It certainly *feels* like we could have done otherwise, and when we observe the behaviour of other people and animals it certainly gives the impression that they, too, could have done otherwise, since we are unaware of anything that would have *compelled* them to act the way they did. At the same time, however, it seems impossible to verify any of this. I can never know whether I or anyone else really could have done otherwise because there is no way to test the hypothesis.[12] In order to do so I would have to be able to go back in time and recreate exactly the same situation, which we know is impossible. Doing something different now than we did last time does not prove anything, because we are no longer in the same situation. The circumstances have changed. So perhaps, despite appearances, we are not free to do otherwise than we in fact do after all.

If you are right that true agency (the kind that both human and non-human animals have but rocks don't) is not compatible with determinism, then it follows either that determinism is false or that there is no such thing as true agency. Yet determinism is not so easily abandoned.

The trouble is that no matter how strongly we feel that we and others have the power to settle certain things, it is difficult to see how this should be possible. It seems that this would require a suspension not only of determinism, but also of our belief in the universal validity of the principle of causality, which is fundamental to our thinking and which makes a commitment to determinism appear inevitable.[13] Whatever happens must have been caused by something, and whatever caused it to happen must have been caused by something else. Nothing comes from nothing. Things are connected by chains of cause and effect. If that were not the case and things could happen that have not been caused by anything, then anything could happen at any time. Yet this, as far as we know, is not the way things work. And even if it were possible for things to happen for no good reason at all, unconnected to and undetermined by anything that happened before, it would not help us to understand the possibility of agency any better. We don't, after all, want to say that what we do hasn't been caused by anything at all. What we want to be able to say is that it has been caused by *us*, which of course begs the question: what has caused us to be the way we are? And if there is something that has caused us to be the way we are (as we must assume), then it seems that we could not in fact have acted otherwise than we in fact did.

*

*Helen Steward*: I agree, of course, that agency does not consist in the power to do just anything. There are things I physically cannot do; there are also many things I could not (as we might say) bring myself to do – and this latter sort of restriction might be just as significant in ruling certain things out for me as are the laws of physics. That is in fact one reason why I think it is very important to characterise the two-way power that is essentially involved in agency not as a power to do A or B, for some specified independent pair of distinct options, but rather as the power, implicit in any active exercise of agency, not to have undertaken that very exercise of agential power.

The 'A or B' way of thinking about things has led some philosophers to argue that when we face decisions which are easy for us, because, for example, the reasons all speak in favour of one option and against the other, nothing goes on which is incompatible with determinism – we are determined by our reasons to do A, say (and would be unable to 'bring ourselves', perhaps, to do the unpalatable B).[14] But this would be a mistaken inference. There may indeed be a sense in which it may be impossible for me to do B, if there is a great deal to be said against doing

it and nothing at all in favour. But this does not mean that my action (when I 'do A') is determined. That I am determined by my reasons on a given occasion doesn't imply that my action (when I undertake it) is deterministically caused. What's determined by the agent's reasons is, for a start, not a token event, but a fact – 'that the agent will (try to) do A' (at some point, in some way, on some occasion – much normally remaining to be settled).[15] And the nature of the determination is not of the brute event-causal sort by means of which the fall of one domino can determine that the next will fall too. It is the agent (and not events inside her) who is said to be determined by her reasons – and what we mean in saying that is merely that the agent formed her intentions on the basis only of the assessment of the relative power of the various reasons at work. Reasons are not causal players which are independent of the agent and her power of action at all – they are merely considerations she responds to in deciding what to do. Their 'power' is metaphorical and utterly dependent upon hers.

Talk of 'determination by reasons' presupposes that an agent with agential powers will be effecting any necessary action. It is not at all the same kind of determining as is implicit in the doctrine of determinism. What determinism would imply is that the actual (token) action I undertake was necessitated to happen, exactly as it did, exactly when it did, etc. And the fact that my reasons for doing something can be overwhelmingly strong does not imply that anything like this is true.

You ask the important question of how we know we are not determined. I agree that in a sense it may be impossible to verify the fact, with respect to any given occasion, that we could have done a different thing. But it is arguable that we are also unable to verify such important and foundational things as the existence of the external world and the existence of other minds.[16] There may be things we have to take for granted in philosophy, despite their unverifiability – and it has always puzzled me a bit that external-world scepticism and what one might call free-will scepticism have been treated so very differently in philosophy (at any rate in recent years). We are permitted to assume the existence of the external world, it seems, despite the fact that no one has really (at least in my estimation) come up with an argument that shows that its existence is more likely than not! But we are *not* permitted likewise to assume the existence of free will (in the sense of alternative agential possibilities). We must worry that determinism might be true, because it is perfectly consistent with all our 'evidence'. But so is the non-existence of the external world, given a certain conception of what that evidence amounts to.[17] In my view, the denial of real, in-the-moment alternate possibilities deals

a blow to our ordinary worldview of much the same sort of magnitude as the denial that the external world exists. Nothing we believe can remain in place if this has to go. If it has to go, there are no agents, there is no (real) thinking, there are no (real) choices or decisions, there is no moral responsibility. There is just 'the dull rattling-off of a chain forged innumerable ages ago'.[18] This is very far from being the world we think we inhabit.

In abandoning determinism it would of course be very problematic if, as you say, we were forced to abandon any principles we ought to take to be obviously true. But I deny that we are forced to do so. You say that we would be forced to abandon the idea that 'whatever happens must have been caused by something'. But there are a number of important points to make about this claim. One is that it appears to be false! There do seem to be spontaneous events in nature (such as radioactive emission events) – we cannot just assume that every event has a cause.[19] A second is that 'caused by' and 'determined by' don't mean the same thing. An agent's doing something may have a variety of causal explanations – and that may enable us to identify a number of things that were the causes of that agent's doing that thing. But that doesn't mean that those causes either individually or jointly determined that the agent's action would occur. A third and final point is that our mental models of how causation works may well not be up to the job of encoding the hugely rich and complex variety of causal (including inter-level) relationships that exist in nature. I've tried to argue, for instance, that it might be possible for a whole to affect its own parts[20] – and that actions might precisely be instances in which whole animals bring about activity in their own sub-systems.

*

*Michael Hauskeller*: It would indeed be a blow to our ordinary worldview if we accepted determinism as true and became convinced that whatever we do is always something we have no choice not to do. Going about our daily business, we naturally assume that many of the things we do we could just as well not do.[21] And when we reflect on our experience, we still feel very strongly that very often we *do* have a choice. We may not, as you say, always have a choice about *what* to do, but even then we still seem to be able to choose *when* to do it and *how* to do it.

We feel equally strongly about the existence of an external world and the existence of other minds, even though the evidence we have in this regard is ultimately inconclusive, so that we can no more prove their existence than we can prove the existence of free will and (true) agency.

So why, you wonder, is scepticism about the existence of the external world or other minds far less common than scepticism about (or even denial of) the existence of free will? It seems to me that this can be easily explained by the fact that while there is very little that militates in favour of the non-existence of an external world and other minds other than our inability to prove their existence beyond any philosophical doubt (which may or may not be *reasonable* doubt), the denial of free will (and hence true agency) follows directly from our commitment to the law of causality and the difficulty of understanding what we actually *mean* by 'free will' or the 'power to do otherwise than one in fact does'.[22]

There is something rationally compelling about determinism that is absent from the denial of an external world or other minds. We may find it difficult – perhaps even impossible – to give up our instinctive belief in the reality of choices, but most people will find it at least equally difficult to abandon their epistemic commitment to the principle of causality. This commitment is in fact so fundamental to how we think about and experience the world that Kant can be excused for concluding that it can only be accounted for if we assume that causality is, just like space and time, not something we happen to come across in our experience of the world, but rather an essential and indeed necessary aspect of the *way* we experience it. Causality, in other words, is not one of the many contents of conscious experience, but part of its very *form*.[23]

It is of course possible that Kant was wrong to assume that the principle of causality is universally valid so that nothing could ever happen without a cause. Perhaps some things do happen without a cause, for instance quantum events or other events occurring at a subatomic level such as the so-called 'spontaneous emission' of electromagnetic radiation. Then again, perhaps they don't and we simply haven't been able (so far) to find the cause.[24] Yet, as noted before, even if we accept the possibility of uncaused events, this does not bring us any closer to understanding the supposed two-way power of agency. Whatever actions are, they cannot be (or be like) spontaneous, entirely uncaused eruptions of a physical nature, because if they were, they wouldn't be actions. For an event to be an action, it must not be uncaused; it must be caused by the agent. And that is the difficulty we face when we try to understand agency. It seems to require something that is in fact conceptually far more difficult to grasp than the suspension of the principle of causality. What we need to get our heads around is not causation or its absence, but the possibility of self-causation. A true agent, it seems, is, like the God of some philosophers: *causa sui* – her or his own cause.[25] Is this something we can really make sense of?

In order to understand agency, we need to carve out a space for the agent to genuinely settle things: things that are neither uncaused nor causally determined by anything but the agent herself. You attempt to create that space by making a distinction between reasons and causes, and also between causation and determination. I want to believe this works, but I'm not sure it does. Reasons are not 'causal players', you say, but merely considerations the agent (freely) responds to. I take this to mean that they don't compel us to act in any particular way. But then, are we free to ignore the relative strength of the reasons that appear to us, so that even though we have most reason to do A, we are still free to exercise our agency by doing B or at any rate *not* doing A? And what about non-human animals, which we believe are also agents? Do they, too, 'consider' reasons for acting in a certain way and reasons for not acting that way? Does talk about reasons and how reasons are different from causes help us understand how animals exercise their agency?

Sometimes, when I am walking my dog and I decide it's time to go home and I call her, she does not come right away but stands still, as if considering whether she should heed my call or ignore it. Then, after a few seconds, she makes up her mind and settles the matter, by either running towards me or staying where she is and continuing to do what she was doing before I called her. I very much doubt that she was engaged in an assessment of the relative weight of reasons to stay or go. So what exactly was going on when, for a few seconds, she was undecided and then, suddenly, she was undecided no longer?

*

*Helen Steward*: I am sure you are right that the reason why scepticism about free will is so much more commonplace than (genuine) scepticism about the external world has something to do with the difficulty we have in understanding how free will can be consistent with something else we think we have reason to believe in – something you call 'the law of causality'. But what does the law of causality say, exactly? Since it rules out free will, it seems important to be very clear what the principle says. As I have said, if it is the claim that every event has a cause, it seems not to be true. Is it perhaps, then, the claim that all *non-random* events have causes? But no one is denying that actions have causes. Among the causes of my getting up may be my realising that it's time to go for my train, for instance. But this doesn't imply that having realised this, I couldn't possibly have done anything else, does it?[26]

Perhaps, then, the thought is rather that all non-random events must have causes which *necessitate* them – causes which are such that no other event could have resulted from them. Perhaps it's tempting to suppose that unless causes are necessitating in this way, there must always be an *element* of randomness about the question whether the result occurs or does not occur. But now we need to know more about what it is for an event to be 'random'. I think it is probably true of events in the inanimate world that if they are not necessitated, then there is something about their occurrence which will have to amount to happenstance – if the event might not have occurred, then however *probable* its occurrence, there still seems to be a chance element involved. But are we allowed to suppose without argument that this principle governs the whole of reality – including the animate world and the actions that it contains? That, it seems to me, is to beg the very question at issue. The question at issue is *precisely* whether it is true that all non-random events must have causes which necessitate them. And it might be argued that actions constitute a major exception to this principle.

We already know, don't we, that any event which is an action is (to that extent) in a sense *not* a random or chance occurrence – for when something is an action, we know already that it was the agent's doing – and that is (in everyday contexts) sufficient to allay the ground-floor question, 'why did *that* happen?'. That is not, indeed, the question we will generally want to ask when we seek to understand a happening in the domain of action: what we will want to ask is normally something of the form 'why did he do *that*?'.[27] The metaphysics is not what puzzles us – what may puzzle us in an everyday context is the agent's motivation. It is not at all like a case in which, for example, a book suddenly falls off a shelf and we wonder what has triggered it. We don't generally have the triggering worry *at all* in connection with things we have identified as actions. Perhaps that is because we already encode actions cognitively in a way that does not presuppose the need for any trigger.

Nevertheless, you may say: as metaphysicians, we *ought* to be puzzled! Can actions just start? How? You say that for an event to be an action, it must be caused by the agent. But I would argue that this is a mistake.[28] If actions are caused by agents, we must ask the question: is there room, or is there not, for the question *how* the agent has caused them? If we say 'yes', we seem to be off on an infinite regress, since the only ordinary way we have of understanding how agents cause things (when they do so intentionally, anyway) is by way of action. A second action would therefore be required to understand the means by which the agent brought about the first one. On the other hand, if we say 'no',

we seem committed to agents causing things without actually *doing* anything – and that seems incoherent. So I think we must resist the idea that agents cause actions. Rather, actions just *are* agents' causings of *other* events – events like arm-raisings and leg-extensions, for example – and also the further effects of these things, like windows being opened and goals being scored.[29]

Still, you may say, actions *begin*. Surely no event in nature can *just begin*! Of course, actions require underlying neurological activity, and to that extent I would agree that we need some sort of account of how the causal principles which govern the underlying neurological hardware are to be rendered consistent with the idea that whether or not an action occurs is up to its agent at the time of action. Ultimately, it seems to me, what we need to understand is not self-causation, exactly, but rather whole–part causation. We need to understand how a whole animal can make something happen in one of its parts – and that requires a better understanding than we currently have of *top-down* causation.[30] I don't think this is easy to come by. But nor do I think there's any a priori reason for thinking it must be impossible. Indeed, I rather think there are a priori reasons for thinking it must be possible! For unless it is, it seems very difficult to understand how the integration of sub-systemic activity which is so obviously a feature of animal life can be maintained. I am now sitting typing at my computer, for example. The muscles which keep me from slumping must be engaged, as must my sense of balance. The knowledge I have of how to move my fingers to hit the right keys at the right time, to type the words I want to write, must be active. I must screen out distracting noises in order to concentrate. And all these things must happen *at the right time* and *in the right order*. Coordination is required. How is it possible to understand such coordination unless wholes are somehow able to influence their parts – to make it the case that the parts are working harmoniously together?

*

*Michael Hauskeller*: Yes, I can see that, but does that really help us understand the possibility and nature of agency? After all, most of those coordinated and integrated sub-systemic activities do not seem to be controlled and directed by me at all. 'I' do not make any decisions about, say, my blood circulation or my digestive processes, and yet they are certainly an essential part of the system that I am, tirelessly working to ensure that system's continued existence. Nor do *I* normally engage my sense of balance or the muscles that keep me from slumping when I sit and write.

These things are clearly coordinated and integrated, but I am not the one who is doing the coordinating and integrating, at least not the I that we are wont to take ourselves to be. I am not even sure it is entirely correct to say that *I* am coordinating the movement of my fingers when I am typing these lines. From my perspective as a conscious, decision-making being, those things just happen, which means they are not actions – or if they are, then I am not really the agent of those actions. There may well be top-down causation, but it is far from obvious who or what is at the top here. The whole may well influence its parts, but which whole or what kind of whole is ultimately in charge? Is my conscious self the 'whole animal' that is keeping it all together? Is my conscious self that which performs that whole animal's actions, that which settles things, or is it just another sub-system that is engaged by that animal to cause certain events? And would it really matter if it were? Do I, or my conscious self, *have* to be in control for those actions to be mine?

Let us look again at what is going on when we make a decision to do something rather than not to do it. I am asleep in bed, the alarm clock rings, I wake up, remember that I have to catch a train, and get up. We can then say that I am getting up *because* I remember that I have to catch a train, or that my remembering this *causes* me to get up. But surely, you say, I could have stayed in bed, *despite* having a good reason to get up. So even if my remembering the train I have to catch *causes* me to get up, it does not *necessitate* my getting up, because I was still free not to. It is hard to disagree with that. However, I don't think anyone, not even the most committed determinist, would claim that a single contributing cause, *on its own*, necessitates a certain outcome. Rather, what is being claimed is that the *whole situation* was such that no other outcome was possible.[31] If I am getting up, and getting up at this particular moment, I am doing this not *solely* because I remember the train I want to catch, but because, in addition, I am sufficiently awake, my bladder feels unpleasantly full, the central heating is already on so that there's no cold room to deter me, I happen to be the kind of person who hates being late, I am hungry and if I don't get up right now I won't able to eat anything before leaving, catching the train is very important to me because I know that if I miss it I won't be able to attend the interview that might get me the job that I desperately want to have, and so on and so forth. There may be a large number of aspects of the given situation that jointly determine what I do, some of which I am consciously aware of and many others that I am completely ignorant of, and none of which would on its own necessitate anything. So strictly speaking a single event is never the determining cause of an effect: it is always the entire state of affairs at a particular

moment in time that *determines* what happens next. Single events can only ever be contributing factors. So if we ask 'why did I get up?' the full answer would have to list all those different factors, which is impossible. This is why we tend to focus on the most obvious, conspicuous or interesting one: because I remembered that I had a train to catch.

Do we now really want to say that despite all these factors coming together to jointly cause me to get up at this moment (and thus to cause my causing), I could *still* have done otherwise? That it was still 'up to me' to get up or not to get up? But was it not up to me anyway (namely to the whole animal that I am)? Did I not *decide* to get up, even though, given the situation I was in, I could not really have done otherwise and it would have made no sense whatsoever if I had? I did not, after all, get up against my will or against my inclinations. Nobody made me get up or manipulated me to do something that I didn't really want to do. On the contrary, getting up is *exactly* what I wanted to do. So why must we insist that I could just as well not have done what I in fact did? What exactly do we *gain* by that?

*

*Helen Steward*: I think there has been some misunderstanding here. I didn't mean to claim that the example I gave of typing, as a case which necessarily involves the integration and coordination of various subsystems, would count as a case of integration and coordination all of which is *done by the agent*. I agree with you that of course much of the coordination which goes on in this case is certainly sub-agential. I only meant to claim that if there were such integration and coordination, it might be an example of something we could reasonably think of as *top-down causation*. And just because *some* such integration and coordination is clearly sub-agential (as you say), it wouldn't follow that *all* of it is.

Once one has had the thought that biological organisms are in general characterised by a hierarchical form of organisation in which top-down (as well, of course, as bottom-up) causation has a role to play, it doesn't seem such an enormous step (at any rate to me) to think that agential control might be, as it were, the topmost form of control in the hierarchy – a way in which a minded organism can respond with *discretion* to rapidly changing environmental information, balancing such things as short-term and long-term priorities, individual gain and social approval, selfish wants and the needs of others, against one another. Do we know that evolution has not found such a discretionary capacity to be an improvement on any deterministic system that might be envisaged

for producing effective and appropriate actions under the complex conditions presented by the ecological niche human beings occupy and the huge part played in that niche by social bonds and social forms of organisation? If so, how do we know? What I dispute is that there is any obvious 'principle of causation' known to be true by us in advance that would rule it out.

I don't in fact disagree with you that there can be circumstances in which the fact that an agent will do some particular kind of thing (φ) in a given situation (e.g. get up soon after their alarm clock has gone off, once they remember that they have a train to catch) is pretty much settled by facts pertaining to that agent's overall constellation of reasons, traits, etc. (together with the absence of certain potential interfering factors, such as physiological impediments). But what I would want to deny is that we can move easily from a recognition of this kind of general settledness and predictability concerning the sorts of things agents will be apt to do under certain conditions to the idea that therefore human agency is consistent with the thesis of universal determinism, from which it follows not only that the *general lineaments* of an agent's doings are determined from time to time by prior circumstances of various kinds, but also that every last detail of timing, trajectory, expression, speed, etc. is thus determined, as the agent moves through the world. That is a much stronger thesis and it requires much more than just the general recognition that (i) what is rational to do is sometimes obvious and (ii) in such cases there is no realistic possibility of the agent's doing otherwise than taking the rational course of action, provided some catastrophic failure of her agency (such as a sudden paralysis) doesn't occur.

The thesis of universal determinism requires that *absolutely everything* is already settled. And it is this totalising deterministic picture that I find very difficult to believe is really consistent with the power to make interventions into the world of which we are truly the source. That there may often be general facts about me, my reasons and my situation which make it more or less impossible to imagine why I would do anything other than φ in a given situation (and hence 'determine' what I do) is something I am absolutely prepared to accept. But the individual actions and activities by means of which I respond to those reasons cannot in my view be themselves the upshot of deterministic processes. It seems important that I have to *execute* those processes, at a time of my choosing and in a way which I control – and these things are essential to the phenomenon of agency. If my so-called 'action' is just produced by neural firings, which in turn were produced by prior neural firings, and those by still earlier events, the whole chain being necessitated, then agency seems to me not

to be what we generally take it to be – a power to *settle* things at the time of action as we move through the world. Like other source incompatibilists, I think this picture turns agency into an illusion.

I agree, then, that we don't *gain* anything from having the power to do things we haven't the least inclination to do. But from my point of view, the question of what we would gain from having this power is the wrong question. That question already assumes that there could be a 'we' even in the absence of any power, on the part of those individuals, to settle things. It assumes that there could be subjects with interests (things to 'gain') who nevertheless lacked the capacity to make events go one way when they really could have gone another. But my question is whether the whole conceptual scheme which surrounds the idea of agency can be made sense of without beings who can settle things, be the source of things. Once such beings are in existence, of course, it is of no *further* help to them to be able to do things they don't want to do, haven't decided to do, can see no reason to do, and so on. But they have to be *in existence* in order to be the sorts of beings to whom things can be of help in the first place – and that is the thing which in my view requires alternative possibilities – although not alternative possibilities of the kind which are usually thought to be required for freedom and/or moral responsibility.

*

*Michael Hauskeller*: I must admit that I find it most difficult to wrap my head around this whole issue. Like you, I cannot bring myself to accept that everything we do, and everything that happens in the world, has been settled all along. Ironically, I feel that I have no choice but to believe in free will – or, more precisely, the power to settle things myself – just as I have no choice but to believe in the existence of an external world and other minds. This is not because the alternative is not conceivable, but because it is part of the fabric of reality as it presents itself to me and generally to beings such as us. The future, I feel very strongly, has not been settled yet. If it had, it wouldn't be the future; it would be merely a kind of past in disguise – a past that we haven't had the chance to become aware of yet. And if the present didn't give us the chance to settle things that hadn't been settled before, then the present would also be merely a version of the past – one that we could observe but not affect in any way. The present, however, appears to be real, which is to say, full of real possibilities, as the past is full of roads not taken that could have been taken when the past was still the present. There are things we can decide, things we can settle ourselves, things that are up to us.

And yet, when I start thinking about it, I am struggling to understand how this is supposed to work and what exactly it means to be the kind of (moderately) free agent that I take myself (and other people, but also non-human animals like my dog) to be. This is because determinism seems rationally compelling to me. Then again, I'm not sure we should worry too much about it. Our discussion of agency has been haunted by the seemingly awful and repulsive spectre of determinism, looming darkly over our ability to act. In a desperate attempt to escape its clutches, we have insisted that it is not real – that it is one of those monsters that reason has a tendency to conjure up. But perhaps we should simply pay less attention to it and understand agency on its own terms. If we did that, we might no longer feel the need to conceive of agency as the power to do otherwise than one in fact does, which is a power we can never be sure we have, nor is it one that it would be particularly desirable for us to have. That is why I asked earlier what we would gain from such a power.

In very simple terms, I exercise my agency when I do what I want to do. This is a power I clearly have. I share this power with other people and other animals. This does not of course mean that it is *always* possible for me to do what I want to do, but very often it is. What I do *not* seem to have is the ability to do what I do *not* want to do, but I do not usually consider this a problem because I have no good reason to want a power that allows me to do what I do not want to do. As long as I am free to do what I want to do, I have all the freedom I need and care about. I do not feel any less free just because I am not free to want what I don't want, or not to want what I want. And if I had that kind of freedom, if I were free to *decide* what I want (which is not the same as figuring out what I want when I am unsure about it), then there would be no *me* that could make a decision. For I am someone, and I am someone because there are things I want to do and things I do not want to do. There are things that interest me, that I care about, that move and engage me in particular ways. If that were not the case, I would be no one. But in order to *do* something, we first need to *be* something. And while we can choose what we do, we cannot choose what we are. Nobody is their own creator.

What and who we are (as a particular kind of biological organism and as individuals) is, to a large extent, not up to us. We find ourselves in the world, shaped physically and mentally by our genes, our history and our environment, none of which is in our control, and we become *someone* through the combined effect of all these forces – a superject rather than a subject. And *then* we act, on that basis: free agents not because we can do otherwise than we in fact do, but because we do what, being what we are, we feel we have to do. Sometimes the clearest expression

of agency is not the arbitrary ('I did this, but could just as well have done that'), but the necessary. Martin Luther, when asked to renounce his teachings by the German emperor Charles V, is supposed to have refused with the words: 'Here I stand. I can do no other.' This is not a declaration of failed agency. It may in fact, on the contrary, be an articulation of agency's very essence.[32]

## Notes

1. For the results of those reflections see my book *A Metaphysics for Freedom*.
2. These different conceptions of free will have generated two different versions of the position generally known as 'incompatibilism' (the view that determinism is incompatible with free will). One is called 'leeway incompatibilism' and argues that *having the capacity to do otherwise* is incompatible with determinism. The other is called 'source incompatibilism' and argues that *being the source of one's own actions* is incompatible with determinism. The most influential recent argument for leeway incompatibilism is that offered by van Inwagen in *An Essay on Free Will*. Roderick Chisholm in 'Human freedom and the self' and Robert Kane in *The Significance of Free Will* both offer arguments for source incompatibilism. For an argument that the two kinds of incompatibilism are in any case connected see Timpe, 'Source incompatibilism and its alternatives' and my 'Frankfurt cases, alternate possibilities and agency as a two-way power'.
3. See in particular my *A Metaphysics for Freedom*, chapters 2 and 3. The notion of settling has some affinities with the STIT ('sees to it that') operator, introduced by Belnap et al. in *Facing the Future*.
4. For a view of this kind see Alvarez and Hyman, 'Agents and their actions'.
5. Why would it be false? It would be false because in a world which contains settlers of matters, those settlers can intervene to prevent and interrupt things that would otherwise occur. See Anscombe, 'Causality and determination' for persuasive arguments.
6. For reflections on the concept of a two-way power see Alvarez, 'Agency and two-way powers'; Frost, 'What could a two-way power be?'; and my 'Agency as a two-way power: a defence'.
7. And not only ordinary language. Notoriously, Actor-Network Theory postulates a fundamental symmetry between human and non-human actors, which includes things as diverse as 'microbes, scallops, rocks, and ships' (Latour, *Reassembling the Social*, 11), and assigns agency to all of them in equal measure. For an introduction see Latour, *Reassembling the Social*; for a critical defence see Sayes, 'Actor-Network Theory and methodology'.
8. There is a school of thought that claims that our sense of agency is in fact deceptive. We *think* we act (freely), when in fact we don't. See for instance the collection of papers in Caruso, *Exploring the Illusion of Free Will and Moral Responsibility*.
9. Compare Midgley, *Beast and Man*, 327: freedom, 'in the sense in which we really value it, does not mean total indeterminacy, still less omnipotence. It means the chance to do *what each of us has it in him to do* – to be oneself, not another person.'
10. For the connection between what we are and how we act, the extent to which our behaviour is predictable, and how this limited predictability affects the dispute between determinism and anti-determinism, see Danto and Morgenbesser, 'Character and free will'.
11. Mary Midgley wrote: 'It is quite easy to be unpredictable, if you don't mind acting crazily. But freedom does not require craziness. Nor does it require omnipotence . . . To be unpredictable, not only to other people but to oneself, is to have lost all control over one's destiny. That is a condition as far from freedom as rolling helplessly downhill' (*Wickedness*, 102–3).
12. See Northcott, 'Free will is not a testable hypothesis'.
13. See for instance Mackie, *The Cement of the Universe*. For an opposing view see for instance Anscombe, 'Causality and determination' and Friedman, 'Analysis of causality in terms of determinism'.
14. For views of this sort see in particular van Inwagen, 'When is the will free?' and Ekstrom, *Free Will*.

15. For a much more detailed discussion of the ontological categories of 'event' and 'fact' see my *The Ontology of Mind*.
16. For further reflections on the parallels between the free will problem and scepticism about the external world see my 'Free will and external reality'.
17. For the classic version of this argument see Descartes, *Meditations* I and II.
18. See James, 'The dilemma of determinism'.
19. See https://en.wikipedia.org/wiki/Radioactive_decay (accessed 1 December 2021) for a brief introduction to the notion of spontaneous radioactive decay.
20. See my *A Metaphysics for Freedom*, chapter 8.
21. And it would be hard, perhaps impossible, not to make that assumption, given how much of our interaction with other people and often also non-human animals relies on it. See Strawson, 'Freedom and resentment'.
22. For a discussion of the various problems the concept of free will faces see for instance Slote, 'Understanding free will'.
23. Kant, *Critique of Pure Reason*, A195–6.
24. For an overview of the current debate on events that appear to have no cause see Svozil's *Physical (A)Causality*, which is dedicated to Kant.
25. See for instance Morden, 'Free will, self-causation, and strange loops' and Strawson, 'Free agents'.
26. See Anscombe's 'Causality and determination' for vigorous arguments against the equation of causality with determination.
27. See Hornsby, 'Action and causal explanation'.
28. Many philosophers have made this argument. For a classic version see e.g. Davidson, 'Agency'.
29. For views along the same lines see Bishop, *Natural Agency* and O'Connor, 'Agent causation'.
30. For discussions of 'top-down' or 'downward' causation see Campbell, 'Downward causation in hierarchically organised biological systems' and the papers collected in Andersen et al., *Downward Causation*.
31. And this whole state of affairs naturally includes not only physical states, but also agential states such as an agent's beliefs and desires. See List, 'Free will, determinism, and the possibility of doing otherwise'.
32. This example is also discussed by Dennett, *Elbow Room*, 133.

## Bibliography


Alvarez, Maria. 'Agency and two-way powers', *Proceedings of the Aristotelian Society* 113 (2013): 101–21.

Alvarez, Maria, and John Hyman. 'Agents and their actions', *Philosophy* 73 (1998): 219–45.

Andersen, Peter Bøgh, Claus Emmeche, Niels Ole Finnemann and Peder Voetmann Christiansen (eds.). *Downward Causation*. Aarhus: Aarhus University Press, 2000.

Anscombe, Elizabeth. 'Causality and determination'. In *Collected Philosophical Papers*, volume 2, 133–47. Oxford: Blackwell, 1981.

Belnap, Nuel, Michael Perloff and Ming Xu. *Facing the Future*. Oxford: Oxford University Press, 2001.

Bishop, John. *Natural Agency*. Cambridge: Cambridge University Press, 1989.

Campbell, Donald T. 'Downward causation in hierarchically organised biological systems'. In *Studies in the Philosophy of Biology: Reduction and related problems*, edited by Francisco Jose Ayala and Theodosius Dobzhansky, 179–86. London/Basingstoke: Macmillan, 1974.

Caruso, Gregg D. (ed.). *Exploring the Illusion of Free Will and Moral Responsibility*. Lanham, MD: Lexington Books, 2013.

Chisholm, Roderick. 'Human freedom and the self'. *The Lindley Lectures*, Department of Philosophy, University of Kansas; reprinted in *Free Will*, edited by Gary Watson, 2nd edition, chapter 1. Oxford: Oxford University Press, 2003.

Danto, Arthur, and Sidney Morgenbesser. 'Character and free will', *The Journal of Philosophy* 54/16 (1957): 493–505.

Davidson, Donald. 'Agency'. In *Essays on Actions and Events*, 43–61. Oxford: Oxford University Press, 1980.

Dennett, Daniel. *Elbow Room*. Cambridge, MA: MIT Press, 1984.

Descartes, René. *Meditations* I and II. In *The Philosophical Writings of Descartes*, volume II, edited by John Cottingham, Robert Stoothoff and Dugald Murdoch. Cambridge: Cambridge University Press, 1984.
Ekstrom, Laura. *Free Will: A philosophical study*. Boulder, CO: Westview Press, 2000.
Friedman, Kenneth S. 'Analysis of causality in terms of determinism', *Mind* 89/356 (1980): 544–64.
Frost, Kim. 'What could a two-way power be?', *Topoi* 39 (2019): 1141–53.
Hornsby, Jennifer. 'Action and causal explanation'. In *Simple-Mindedness*, 129–53. Cambridge, MA: Harvard University Press, 1997.
James, William. 'The dilemma of determinism'. In *Essays in Pragmatism*, 37–64. New York: Hafner, 1970.
Kane, Robert. *The Significance of Free Will*. Oxford: Oxford University Press, 1996.
Kant, Immanuel. *Critique of Pure Reason*. Translated and edited by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press, 1997.
Latour, Bruno. *Reassembling the Social: An introduction to Actor-Network Theory*. Oxford: Oxford University Press, 2007.
List, Christian. 'Free will, determinism, and the possibility of doing otherwise', *Noûs* 48/1 (2014): 156–78.
Mackie, John Leslie. *The Cement of the Universe: A study of causation*. Oxford: Clarendon Press, 1974.
Midgley, Mary. *Beast and Man: The roots of human nature*. Hassocks: Harvester Press, 1978.
Midgley, Mary. *Wickedness: A philosophical essay*. London: Routledge & Kegan Paul, 1984.
Morden, Michael. 'Free will, self-causation, and strange loops', *Australasian Journal of Philosophy* 68/1 (1988): 59–73.
Northcott, Robert. 'Free will is not a testable hypothesis', *Erkenntnis* 84/3 (2019): 617–31.
O'Connor, Timothy. 'Agent causation'. In *Agents, Causes and Events: Essays on indeterminism and free will*, edited by Timothy O'Connor, 173–200. New York: Oxford University Press, 1995.
Sayes, Edwin. 'Actor-Network Theory and methodology: just what does it mean to say that nonhumans have agency?', *Social Studies of Science* 44/1 (2004): 134–49.
Slote, Michael. 'Understanding free will', *The Journal of Philosophy* 77/3 (1980): 136–51.
Steward, Helen. *The Ontology of Mind: Events, processes and states*. Oxford: Oxford University Press, 1997.
Steward, Helen. *A Metaphysics for Freedom*. Oxford: Oxford University Press, 2012.
Steward, Helen. 'Free will and external reality: two scepticisms compared', *Proceedings of the Aristotelian Society* 118/2 (2019–20): 1–20.
Steward, Helen. 'Agency as a two-way power: a defence', *The Monist* 103/3 (2020): 342–55.
Steward, Helen. 'Frankfurt cases, alternate possibilities and agency as a two-way power'. In *Fifty Years of Responsibility without Alternate Possibilities*, edited by Geert Keil and Romy Jaster. Special Issue of *Inquiry* 2021: https://doi.org/10.1080/0020174X.2021.1904639.
Strawson, Galen. 'Free agents', *Philosophical Topics* 32 (2004): 371–402.
Strawson, Peter F. 'Freedom and resentment', *Proceedings of the British Academy* 48 (1962): 1–25.
Svozil, Karl. *Physical (A)Causality: Determinism, randomness and uncaused events*. Cham: Springer, 2018.
Timpe, Kevin. 'Source incompatibilism and its alternatives', *American Philosophical Quarterly* 44/2 (2007): 143–55.
van Inwagen, Peter. *An Essay on Free Will*. Oxford: Clarendon Press, 1983.
van Inwagen, Peter. 'When is the will free?', *Philosophical Perspectives* 3 (1989): 399–422.