# Top-Down Causation in the Animal Kingdom

By James Clark Ross

Supervisor: Dr Ben Springett

Department of Philosophy — Birkbeck, University of London

**Abstract:** The problem of free will has roots in Ancient Greece. Yet the question of whether humans are truly free or not is far from being resolved. Helen Steward looks to the animal kingdom for rudimentary signs of freedom, arguing that animals which possess the capacity to bring about bodily movements in purposive ways are 'animal agents'. Steward claims animal agents do this in a way which is irreducible to their bodies' physical properties as part of top-down causality. In this paper I put forward two objections to her view: (i) Steward cannot claim that animal agency is a real phenomenon without knowing that the thesis of determinism is false; (ii) Steward's conception of animal agency is unintelligible. I argue that, while her project survives (i) on the grounds that we ought to be agnostic with respect to determinism, it falters on (ii) on account of begging the question in favour of top-down causality.

**Keywords:** agency, causation, free will, metaphysics, philosophy of science

*Thank you, Ben Springett, for your invaluable supervision.*

*Thank you, Dijana Vilić, for your unrelenting support.*

# Table of Contents

# 1. Introduction

## 1.1. The problem of free will

Do we ever act freely? 'Free will' denotes the ability to make intentional decisions at one's own discretion. It is inextricably bound up with a myriad of human activities such as 'deliberation', 'rationality', and 'creativity'. These activities presuppose the truth of free will, where without it they become mere descriptions of conscious processes, not instruments used by free-thinking agents to reach goals on behalf of themselves. Freedom's absence should force us to reconsider our ascription of moral responsibility, for people ought not be held accountable for actions they did not freely perform, calling into question the notion of retributive justice (Pereboom 2013).

The problem of free will is understanding how we are free to make decisions when they might be determined for us. If we believe that events in the Universe unfold naturally according to physical laws, we must discern the origins of our 'intentional decisions' from these physical laws.

The primary issue I take with much of the free-will discourse lies in a difficulty to foresee significant epistemic progress from it: unknowns and complexities—with respect to mind, self, and possibility—stall many projects, keeping free will a mystery of metaphysics which has profound but unknown implications. We should be careful not to disregard the debate if we care to build a just society in which moral responsibility is fairly ascribed to individuals. The difficulty, however, as per the central themes of this paper, is located in understanding the *causal role of agents* in their actions. If we can understand this, we can determine how much agency, if any, they had in those actions and, accordingly, how much moral responsibility they should be ascribed.

I argue that if we seek epistemic progress from our research, the free-will debate should be reformed to search for freedom in rudimentary activities of agents, not in complex activities in the context of moral responsibility. This is because, c*eteris paribus*, more-basic philosophical claims can be assessed in more-rigorous detail. In light of this stance, I have sought to dissect a contemporary theory of freedom which is centred on *simpler* claims. Specifically, I have focused on Helen Steward's *A Metaphysics for Freedom* (Steward 2012). Steward looks to the animal kingdom for a rudimentary mode of freedom: namely, physical action through one's bodily movements. Steward does not promise a complete theory of free will: she only hopes to expose what Jean-Paul Sartre (1958) called the metaphysical 'structures' of action (Steward 2014). In my view this methodological isolation may better facilitate epistemic progress with respect to the problem of free will, for pursuits of more-comprehensive conceptions of freedom can be built on the same structures. To introduce human concepts such as morality into the debate is to dive straight into the deep end. While this would appeal to our interests, the metaphysical structures of action by themselves are sufficiently challenging to delineate.

### 1.2. Key concepts

To make sense of Steward's work in the context of the greater debate I have explicated the following free-will concepts in Stewardian terminology: agency (§1.2.1), causation (§1.2.2), determinism (§1.2.3), indeterminism (§1.2.4), and agent causationism (§1.2.5). Their definitions are precursors to hopefully a clear exposition of her work (§2).

### 1.2.1. Agency

Freedom is a concept which free-will metaphysicians attempt to delineate (or refute), notwithstanding disagreement on its precise meaning. Steward argues that a variety of freedom is expressed in the *agency* of certain animals, which denotes the capacity to internally bring about bodily movements in the external world. They express freedom because *they* decide how to move *their* body, rendering the future open according to the actions that follow. Their options are constrained, naturally. For instance, the movements of a predatory cat might lead a moth to fly away. But even though the moth flew away in response to the *cat's* movements, it was the moth which (hypothetically) brought about its bodily movements and could have flown away in a number of ways (or not). As Steward states:

> 'A deer is clearly not free not to run from a lion it has spotted running towards it, a spider not free not to bother with spinning any webs for a few weeks. It is utterly undeniable that all animal agency takes place within a framework which constrains, sometimes very tightly, what can be conceived of as a real option for that animal…What I wish to insist upon is only that there is much flexibility within these constraints, even for very simple creatures, for such things as different orderings of the actions necessary to complete a complex task or set of such tasks, the taking of alternative spatial routes to a place, different chosen strategies for achieving a given goal, different timings…Doubtless, the precise degree of flexibility which is possible depends upon the sophistication of the animal.' (Steward 2012, p. 20)

Steward's conception, on the face of it, is simple: an agent possesses 'the capacity to move oneself about the world in purposive ways, ways that are in at least some respects up to oneself' (Steward 2012, p. 4). However, upon further scrutiny, it becomes apparent that what constitutes 'capacity to move oneself' and 'purposive' is rigorously

demanding. To counter this criticism she introduces the process of 'settling' (§2.3), which I have subsequently challenged the intelligibility of (§3.2).

Steward focuses on agential powers which are remarkable enough to warrant significant consideration but not remarkable enough to be unique to human agency. However, the idea that agents make decisions originally to move their bodies disrupts how we usually construe causation in nature (§1.2.5).

### 1.2.2. Causation

Causation is generally a description of how past states (causes) progress to future states (effects), whereby certain future states arise *because of* certain past states. The unknown of what connects the two embodies the problem. It is not sufficient to say that causation is self-evident by virtue of one coming before the other. A young person becomes an old person: arguably, no set of causes is at play—just a series of events.

Causation involving agents and their movements could be illusory. An empiricist, on one common reading of Hume (2008), might claim that we do not possess the knowledge to control events in the external world at all; that we only project our expectations, born from personal experiences, onto them. Steward's metaphysical structures of action, which delineate a causal framework of movement, contradict this idea with a notion that agents bring about movements that actively change states of things.

Causation, then, is not strictly a metaphysical concept that pertains to animal agents, for it could apply to any set of stative relations between objects (e.g. rainfall and plant growth). However, in this paper I am only concerned with the role of animals in causality and how their movements result from a metaphysical process distinct from

physical relations. Steward posits that agents freely move their bodies on account of top-down causation—that is, by means of decisions sourced in agents' minds which are not reducible to *sub-agential* phenomena (such as biochemical events).[1] She argues that such causation involving agents and their movements is incompatible with determinism (§2.2).

### 1.2.3. Determinism

Determinism is the thesis that events are determined by previously existing causes (e.g. the Universe's basic physical properties or some preordained laws set by a supernatural deity). If the Universe is entirely deterministic ('universal determinism'), we might suppose that agents cannot exist because agents' actions are like everything else, determined, only ever entailing a single physically possible future and nullifying the causal role an agent can play with respect to settling one of multiple physically possible futures (§2.3). Steward is only concerned with refuting a more-localised variant of determinism's thesis, according to which agency could be superimposable over deterministic, sub-agential phenomena. This possibility would be enough to pose significant problems for Steward since she requires agents to be in charge of their movements, though some philosophers, 'compatibilists', still support the idea that determinism does not preclude agency.

[1] Steward uses 'sub-personal phenomena'. To clarify that her views pertain to all *animal* agents I have used 'sub-agential phenomena' instead.

### 1.2.4. Indeterminism

Indeterminism is the thesis that not all events are determined by previously existing causes. Steward argues that, for animals to be agents, indeterminism must be obtained at the level of agency. This puts her into the group of philosophers called 'incompatibilists', who see determinism as incompatible with agency.

The prevalent view amongst physicists is that indeterminism reigns at the Universe's fundamental scale, for quantum mechanics characterises indeterminism with great accuracy.[2] However, Steward is aware that quantum mechanics might not manifest in our thoughts in ways meaningful to agency because its workings cannot be applied to systems so large in scale (Honderich 1988; Weatherford 1991). Steward also accepts that the absence of determinism by itself does not explain how agents can be connected to a microphysical reality such that they are involved in causation through their bodily movements. So Steward looks to agent causationism to help solve the problem of free will.

### 1.2.5. Agent causationism

Steward's philosophy amounts to an agent-causal view. In accordance with agent causationism, *agents* are able to start new causal chains. Steward's central agent-causal claim is that animal agents bring about movements in their bodies without prior cause

[2] An electron enclosed in an atom, for example, is treated as a wave whose non-locality are represented by a continuous probability distribution, while the time of radioactive decay can never be precisely predicted. Do phenomena involving subatomic particles somehow propagate upwards to agency?

to initiate action. This position sits in dialectic opposition to event causationism, which conceptually brings together agents and causation in virtue of *events* involving both.[3]

Steward's theory might raise some eyebrows in the scientific community, though, for scientists may claim that all physical events, in bottom-up fashion, are explainable through physical relations, even though we have not proven this *yet*. In such a worldview it would be true that nothing is *causa sui* (original): all events would be causally connected to antecedent events, leaving no room for free action (Strawson 1994). Correspondingly, scientific explanation would see animals as merely complex physical objects, not agents, that operate according to the laws of nature.

Steward must now convince us *how* agents are metaphysically constituted. She asks us to believe that animals can be the 'true authors' of their actions when they bring about self-movement. However, causation, in line with this explanation, is irreducible to sub-agential phenomena, which undermines a frequently reductive view of causation in science (as understood by physical laws). Despite this, Steward's agent causationism need not represent a superfluous commitment to supernaturality since she only asks us to reconsider a reductive understanding of causation. We ought to be sceptical of reductionism in agency, for it underlies but one of many possible descriptions of causation.

[3] I do not have the room to provide a full breakdown of event causationism. Nonetheless, to illustrate its basic principles consider this saliently put example. Imagine a capuchin monkey contemplating deshelling a tough, edible nut. The capuchin monkey's agency, it might be claimed (Kane 2002), is located in its conflicting motives, where one motive is to crack the nut for its calories and another is not to inefficiently expend energy doing so. The final decision is random with respect to mental states that transpire within the capuchin monkey. However, because the capuchin monkey's mental states led to the generation of these conflicting motives in the first place, its agency is located between these mental states and the eventual outcome.

## 2. Exposition

Up to this point I have described Steward's view in general terms. In this section I have provided a detailed breakdown, introducing and scrutinising her main claims. First, I have summarised her strategy (§2.1). Then I have described what I see as her most-crucial assertions (§§2.2–2.4). Following on from this dissection of her work I have posed some specific obstacles to the progress of her project by critically evaluating her central claim that animals can bring about bodily movements without any causal relations to antecedent events to initiate action (§3).

### 2.1. Steward's strategy

Steward's work can be rendered into a two-pronged position: her first aim is to derail compatibilist approaches to freedom; her second aim is to deliver an incompatibilist theory of freedom. With respect to her first aim, Steward expounds 'Agency Incompatibilism' (§2.2). With respect to her second aim, Steward constructs a sufficiently broad and simple account of agency that can assign freedom, in rudimentary form, to various animals across the animal kingdom. She does this by appealing to some key metaphysical concepts (§§2.3–2.4).

In espousing an evolutionary picture of the natural world Steward argues that humans, in their dispositions, are probably continuous with non-human animals. The corollary of this, if she succeeds at accommodating a safe definition of animal agency into this evolutionary picture, is that human freedom can be claimed too, for we are animals as well.

### 2.2. Agency Incompatibilism

Steward uses 'Agency Incompatibilism' to denote a kind of compatibilism which applies to agency specifically as opposed to freedom generally. However, the charge at compatibilism is the same: freedom, as displayed in animal agency (if there is such a thing), is undermined by determinism.

To understand Agency Incompatibilism we should look to Steward's somewhat-radical notion of agents. On her view, action, in any meaningful sense of the word, requires an agent that possesses a certain 'set of powers' to settle one of multiple physically possible futures (§2.3). These powers are irreconcilable with universal determinism because determinism roots action to antecedent events (i.e. as predetermined neural events, muscle contractions, etc.). While many bodily events *are* rooted in antecedent events like this and are required by animals to move, they do not constitute *action*. To explain how agents bring about movement outside of enchained physical relations Steward appeals to a pluralistic notion of causation, which I have explicated later (§3.2.1).

Following Steward's line of thought, proving that the thesis of determinism is true would rule out agency and free will altogether since a deterministic universe only has a single physically possible future. Steward, however, doubts that physics—probably our most-fundamental scientific discipline since it involves exploring the behaviour of the Universe at fundamental scales, from particle interactions to the expansion of space—will *necessarily* settle the debate. We have reason to doubt the efficaciousness of the scientific worldview altogether: the physical laws that scientists come up with might only apply to a specific set of situations within a single domain of relations (i.e. the domain of physics for which the laws were derived). It is controversial to say we

should doubt physics' suitability to settle the debate when physics continues to sporadically revolutionise our understanding of the Universe. I have addressed this concern later (§3.1).

In the opinion of philosopher of science Nancy Cartwright (1999), who opposes a reductive pyramid of explanations (Figure 1), our understanding in any domain of knowledge (e.g. human behaviour according to neuroscience) cannot be conferred from lower domains (e.g. laws of physics at the pyramid's base).

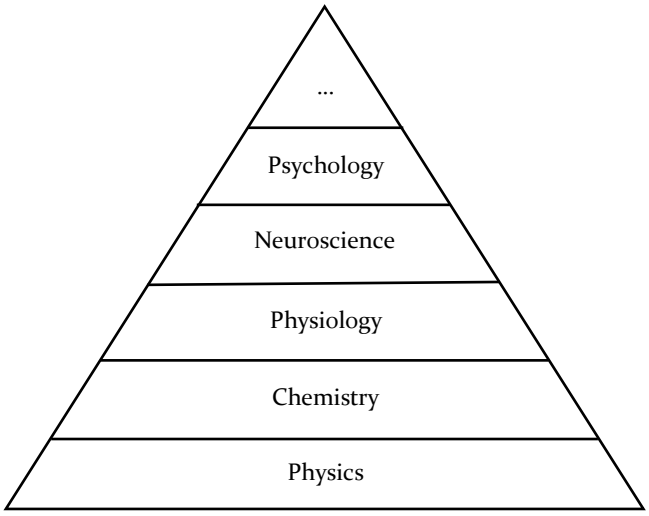


**Figure 1:** A reductive pyramid of explanation. There are separate domains of relations at its different tiers, within which domain-specific relations between events can be postulated and attested by empirical findings. These relations, in theory, can be employed to understand relations in different domains. This picture threatens irreducible agency, for agency's explanation could come from one or more of these domains as opposed to an irreducible metaphysical domain.

With the threat of determinism held at bay on this account, we must now ask Steward how agency is constituted if it is not reducible to deterministic, physical laws. One possibility, according to Steward, is that agency did not evolve from a *physical* reality, meaning we cannot rule out separate, metaphysical origins (e.g. a deity).

Another possibility, which Steward focuses on and which I have scrutinised accordingly, is that agency has evolutionary roots. This claim is compatible, so to speak, with our scientific understanding of evolution in the sense that animals, on one hand, can be explained in terms of evolutionary biology (e.g. DNA, physiological traits, etc.); whilst, on the other, agency can belong to a different, non-physical ontological category altogether. I have expounded and analysed these claims later (§3.2).

The subsequent two sub-sections (§§2.3–2.4) reveal two crucial concepts of Steward's theory that together show us how Steward believes the future is open according to self-movement by agents.

### 2.3. Settling

Agency, according to Steward, hinges on one's ability to settle.[4] Through this ability animals frequently decide to internally bring about certain bodily movements (e.g. extend a tongue) to initiate an action that physically changes the states of things in the external world (e.g. catch a fly). Agents possess the following abilities, however rudimentary, to settle: move their bodies (§2.4), make a decision born outside deterministic chains of events, be a centre of subjectivity, and exhibit intentional states.

I should first crystallise the relation between agents and the space-time they purportedly move their bodies through in our definition of settling. When an agent decides how to physically move their body they settle matters (i.e. one of multiple physically possible futures). A matter can only be resolved by an agent; the matter is

[4] Steward refrains from using the term 'determine' here. While it would suit the explanation, she seeks to avoid etymological association with 'determinism' in expounding her incompatibilist take on freedom. I have continued in this vein.

unresolved—and, therefore, indeterminate—up until that spatiotemporal point, which is when the action ensues.

Steward postulates that settling only applies to matters that are settled in advance in physically *possible* ways at the time of settling. For instance, I can only settle the action of running along the entirety of Earth's Equator in 24 hours if that claim can somehow be spatiotemporally fulfilled. Clearly, I do not boast the capacity to meet such a claim.

Settling must also be *up to* agents such that they have a part to play in causation. That is, decisions have to bear indeterministic relations with antecedent events *and* be up to agents (Figure 2); else 'actions' are just *indeterminate events*. Such 'up-to-usness' gives animals' decisions purposiveness in line with the following definition of settling:

Settling: *An agent settles a matter iff* they *decide to* ϕ *at time* T *in way* W *at place* P.

That is to say, agents settle matters on these terms when they decide to bring about certain movements. A challenge comes from a claim that animals' decisions, though subjectively held and made with intention prior, are fundamentally governed by natural laws still (e.g. quantum mechanics propagated to the domain of agency). Animals' considerations prior to movements may feel unique to them, in line with their specific intentions. However, it does not necessarily follow that they *freely* came up with them. A settled matter could just be a culmination of probabilistically unfolding of events in a number of unpredictable ways, which would apply to conscious considerations, undermining animals' agency by negating their role in causation through a known physical relation between their considerations and the Universe's basic physical

properties. Steward deals with this challenge by claiming that agents *are* the true authors of their actions. Settling is fundamentally important to *what* movements come about: there are multiple physically possible futures and the agent settles which one comes about.

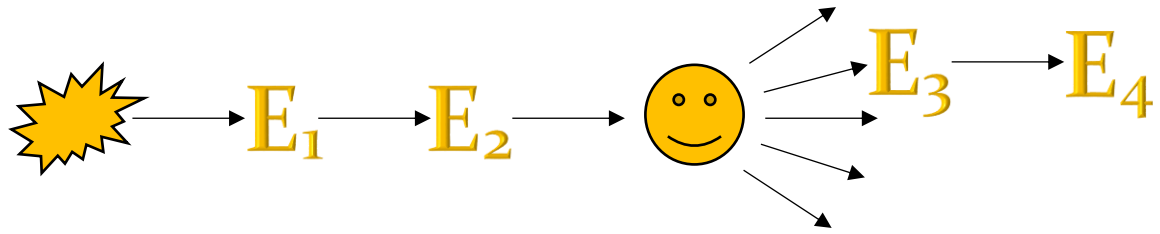


**Figure 2:** Agents bear indeterministic relations with respect to antecedent events; else the Universe, from the Big Bang onwards, would simply be a series of events, leaving no conceptual space for causation.

Therefore, Steward's view invokes an indeterministic notion of the Principle of Alternative Possibilities (PAP), which I have expressed as follows:

PAP: *An agent is only free if they could have done otherwise before an open future.*

Strictly speaking, Steward does not require robust PAP: animals only need to be able to perform actions under a non-compulsory exercise of power. She calls this ability a 'relevant refrainment power'. At a minimum, this grants to animal agents an option to settle a matter *and* an option not to settle it. For example, a squirrel displays agency if it can jump to a tree branch or refrain from doing so, assuming both movements are physically possible. On the other hand, a compatibilist may challenge this claim by

arguing that the squirrel, enchained to one physically possible set of body movements in the future (e.g. not jumping), can choose to do otherwise in the immediate run-up to the event of jumping *if it wants to*. As such, one might claim that PAP is not required for an agent can be a *source* of intention (Frankfurt 1969). Is this enough for agency? Steward would contest this by saying that for an animal to be free they must face multiple physically possible futures such that their intention is not determined for them. This gives agents room, conceptually speaking, to freely decide future states in the Universe from current states in line with our notion of causation (§1.2.2). At any rate, PAP, in Steward's mind, is incongruous with determinism since settling always originates in agents in non-mechanistic, top-down fashion. Steward can also ask the compatibilist: 'What gives determined animals intentions?' It becomes hard for them to provide an answer without an infinite regress of explanation since a 'decision' would arguably be an event, $E_n$, entailed in a long chain of other events, $E_1 \ldots E_n$.

Here is a good point in the exposition to ask Steward how the relations of settling are discernible from the sub-agential properties of agents' physical bodies in commanding their movements. Her view relies on a notion of substance causationism to explain an 'owner-body distinction', whereby the agent can only be treated as such when considered to be a 'substance': an ontologically fundamental entity which boasts the required properties of agency in virtue of interdependencies between events (§3.2.1). Agency, then, is not reducible to sub-agential properties but constituted *qua* whole substance. While we might attempt to explain an agent's thoughts by decomposing them, for example, into measurable flashes of brain activity, or by understanding them through the agent's dispositions and experiences (as per folk psychology), either would presuppose a form of reductionism that Steward's non-mechanistic view does not adopt.

This set of postulations faces challenges. Firstly, if matters have to be up to agents, in what way, *exactly*, are agents irreducible with respect to the bodies they make decisions for? She needs to intelligibly demarcate the metaphysical structures of action to show that they do not fit within a deterministic interpretation of agency (§3.2). Secondly, a compatibilist might argue that agency is neatly superimposable over deterministic, sub-agential phenomena—for example, as squirrel's intentions *atop* their determined jumping or not jumping.[5] As per Daniel Dennett's influential 'Intentional stance' (Dennett 1971), if an animal (an intentional system) possesses the ability to act on mental activities, such as beliefs and desires, we can predict its behaviours based on what we think it will do as a creature that intends to further its goals based on those beliefs and desires. As such, we can find what the incompatibilist is seeking at the macroscopic level of intentions and wed it to the deterministic, 'non-Intentional' domain of physical sciences and its predictive power. Thus we can debatably meet three of Steward's criteria of agency (Broadie 2013): namely, an agent can move the 'whole, or at least some parts, of something we are inclined to think of as its body', be 'a centre of some subjectivity', and be attributed 'at least some rudimentary types of intentional state (e.g. trying, wanting, perceiving)' (Steward 2012, pp. 71–72).

Steward, however, is at odds with such compatibilism. She argues that something like the intentional stance contradicts meaningful action since agents would not genuinely face open futures (as per PAP), leaving nothing up to agents to *effect causation*. Furthermore, she makes a charge at Dennett's creeping 'gradualism' since one can apply an intentional strategy to almost any object. Correspondingly, Dennett

[5] This view might, however, have serious moral implications. It would be unjust to ascribe moral responsibility for the bottom-up instantiations of things. For what can 'agents' do about the properties of things?

does not refuse to apply it to inanimate objects such as thermostats, whose temperature-stabilising behaviours can be predicted if we attribute an ability to hold beliefs about the temperature to it. Steward puts it as follows: in Dennett's mind, consciousness 'imperceptibly fades into unconsciousness, mindedness into mechanism' (Steward 2012, p. 113). However, Steward posits that agents' subjectively born mental states play a role in deciding future bodily movements, which requires consciousness, however rudimentary, to form intentions.

With respect to the semantics of action, Steward borrows the Anscombe–Davidson approach (Anscombe 1957; Davidson 1971). Accordingly, Steward states that when 'an agent $\varphi$-s by $\psi$-ing, her $\varphi$-ing may normally be identified with her $\psi$-ing' (2012, p. 34). *Viz*, for each action there is a number of other descriptions it can equate to. If we take $\varphi$-ing to mean a cheetah 'drinking water' as an example, we could identify $\varphi$-ing with 'drinking water', 'bowing head down to water and making successive tongue movements', or similar for '$\psi$-ing'. There are two issues here that might foster some worries which are worth noting but not fully calling into question, given the scope of this paper. Firstly, Steward's actions can be sub-classified into successful and unsuccessful actions.[6] Does a single definition of settling apply equally to both? Secondly, in Steward's adaption of the Anscombe–Davidson approach she incorporates sub-intentional actions (O'Shaughnessy 1980). This may foster more worries. Her adaption is intuitive in some sense: if I feel an itch as I am mesmerised by the sentences I am typing out on my laptop, I might absentmindedly decide to scratch it; and Steward wants to say that *I* settled scratching my leg by deciding to move my body,

[6] And sub-classifications of agents, potentially breaking down the unity of agents, as I have gone on to describe with regard to 'immanent' events later (§§3.2.1–3.2.2).

notwithstanding my being on autopilot. Steward does not seem to be too concerned but I am more so. By diluting 'intention' like this she blurs the boundary between the actions of agents and regular events entailing simple biological entities which are incapable of conscious intention. Yet she seeks to subsume countless animals into a single category—that of animal agents. Arguably, then, Steward falls into the same gradualism as Dennett. However, Steward appreciates that animals' minds vastly range in sophistication but has no room in her work to explore which animals—from lowly entities, such as clams and earthworms, to cognitively advanced entities, such as orangutans and elephants—are categorically agents.

In summary, settling is a broad metaphysical concept whose structures might apply to action. True agency, according to Steward, is found in possessing the ability to settle which actions to perform out of one of multiple physically possible futures. While an agent needs lower-level properties to act, their agency is found, irreducibly, in settling actions from top to bottom *qua* whole substance. This, in Steward's mind, is what creates an owner-body distinction, a discussion of which comes to the fore again later (§3.2).

### 2.4. Self-moving

Animal agents move at least some parts of their bodies. However, animal agents do not *cause*: the bodily movements they bring about in settling form a *component* of top-down causality (§3.2). Of course, not all bodily movements (e.g. contractions of cardiac muscle) can be settled. Meanwhile, many biological entities (e.g. sperm cells) are seen to move, whilst others do not move at all (e.g. sessile sponges). Neither kinds of entity

seem to bear any powers of settling—say, the ability to make a decision, to plan, or to intend—and should not be understood to exhibit agency.

Steward puts 'self-moving' and 'open future' together in the following valid argument, where *S* is 'Agency consists in bringing about self-movement' and *O* is 'The future is open':

$$S, S \rightarrow O \therefore O$$

This argument aligns with what has already been said about Agency Incompatibilism. Buttressed by the notion of settling, Steward asserts the truth of *S* by way of arguing agency as an ability to settle matters. Therefore, if she corroborates *S*, *O* is a valid conclusion (rendering the thesis of determinism false). But what exactly constitutes self-movement?[7] What significant, metaphysical steps are there between, say, animals and physical objects that move 'themselves' (e.g. timed sprinklers and robots)? Well, to clarify, while inanimate objects move, in some sense, on their own accord, they do not self-move: agents remotely program them to move in purposive ways by settling their *own* movements. The intelligibility of Steward's answers have been put to the test later (§3.2).

[7] As a point of semantics, there are some important nuances to what we mean by 'self-movement'. In the context of agency we are concerned with self-movement of the body as an action but this is not meant to include *all* bodily movements. Agency can only be found in *transitive* movements—the doings of things, not the results (*intransitive* movements). Ambiguity between the two kinds of words, however, arises in the English language. Take the event of chocolate melting: 'melting of the chocolate' could refer to chocolate melting because this is something which merely happened to the chocolate and it could refer to a person being proactive in melting it. In the same vein there is a semantic difference between something moving as an action and as a result. We are concerned with the former.

# 3. Critique

In this section I have offered two objections to Steward's central claim that animal agents bring about bodily movements without causal relations to antecedent events. My challenges are made with respect to the following possibilities:

(i) The thesis of determinism is true.

(ii) Her theory is unintelligible.

In exploring how Steward might retort I reveal how she might overcome these blows in some respects.

## 3.1. The Epistemological Argument

According to Steward, there are no freedoms worth wanting if the thesis of determinism is true at the level of agency. Determinism precludes action, in her sense of the word, since there would be no entities that settle matters to initiate action. Matters, instead, would have been settled already by the initial conditions of the Universe and the basic physical properties that followed. The truth of determinism, then, would completely undermine Steward's conception of agency, for she would need to concede that agents did not exist, as expressed in standard form below. However, the specific problem for Steward is that she argues that *there is such a thing as agency without knowing whether the thesis of determinism is true* or *false*. Both determinism and indeterminism are open epistemic possibilities yet she hedges her bets on the thesis of determinism being false to claim that there *can be* such a thing as agency.

**P1:** Determinism and agency are incompatible

**P2:** The thesis of determinism is true

---

**C:** There are no agents

### 3.1.1. The argument

Because this attack is made with respect to our knowledge concerning agency, Steward calls this challenge the 'Epistemological Argument'.

Before I explore what routes Steward has access to here let me eschew a certain take on agency. That is to say, if the thesis of determinism was somehow demonstrably true, Steward would be forced into concluding that agents did not and could not exist. But is this not absurd? It is obvious that they exist; that many animals clearly have multiple futures open to them. It is easy to observe, for example, deliberation (if we can call it that) in a hungry wood mouse as it 'contemplates' whether to fetch seeds or stay safe within its nest. Maybe, then, we can say that agency is *self-evident*; that the problem of agency is only situated in *explaining* how. So can C be refuted like this? To the contrary, I argue that while some facts might not need to be explained (§3.2.3), the notion of animal movement at least ought to be put under the microscope such that it can be explained with respect to those facts. By assuming agency we carelessly do not capture *how* something like a wood mouse contributes to causation.

Meanwhile, we should look for ways through which Steward can escape the Epistemological Argument whilst upholding her argument against it. Steward shows clarity with regard to P1 through Agency Incompatibilism. However, now we require justification as to why she can skip the truth-value of P2, given that she refutes C.

If Steward tries to convince us that agents exist (to refute C) through an independently constructed premise, her argument will become dangerously close to falling into Moore's paradox (Moore 1993). That is, she would be making a proposition, P1, whilst holding no unequivocal knowledge on the subject of determinism, with regard to P2. While it is *logically* consistent to do so, it is counterintuitive for the two to be strung together in one argument. It would be analogous to the statement: 'God gave Moses tablets of stone etched with the Ten Commandments but I do not know whether God exists.'

Alternatively—and this is her own trick—Steward expressly returns the burden of argument to her challengers. Steward does not deem the Epistemological Argument to hold sufficient clout to undermine her entire argument. An opponent merely asserting that it is would amount to question-begging: *they* need to justify their claim. On Steward's own account the thesis of determinism is false; thus she disputes P2 to refute C. She is, however, happy to admit that it is *conceivable* that she will eventually be proven wrong on P2. But, she asks, why should we take this as a real contest before a real qualification is made?

Nevertheless, a proponent of an entirely scientific view of the Universe could oust her of agency before her theory took flight by claiming that only a scientific explanation can settle P2. Correspondingly, she would be unable to argue that the thesis of determinism is false through *philosophy*. The challenge is a job for *science* and she should wait and stop speculating. The scientific worldview has allowed us to understand so much. Why take agency out of its remit?

However, scientists can only claim, with scientific integrity, to understand what we have empirically demonstrated *so far*. Thus a contention that a scientific

understanding ought to rule on determinism would represent an uncharacteristic leap of faith. So, in the face of such contingency, we ought to remain agnostic with respect to its role in proving or disproving the truth of the thesis of determinism and, by extension, animal agency.

Determinism may or may not underpin agency. Hence Steward's view cannot be dismissed through the Epistemological Argument. It amounts to a weak challenge through question-begging. The facts concerning determinism may eventually deal a fatal below to Steward's view but they, as of yet, have not been realised, keeping her theory alive.

### 3.1.2. The scientific method

Now I want to go further. I want to suppose that scientific method is considerably more stunted than usually thought when it comes to revealing the true workings of the Universe. That is to say, even if scientific knowledge *appears* to strike animal agency down at one point in time, there would still be room for Steward's arguments to hold truth. Whilst being open to the implications of scientific knowledge this pre-emptive move would add question marks around it, thus further preventing the Epistemological Argument from collapsing Steward's theory.

Not only do we lack conclusive, empirically found evidence of determinism at the level of animal agency but our scientific understanding of the world may *never* undermine animal agency. It is only a *possibility* that future empirical findings will force us to accept determinism. Scientists, therefore, have no exclusive right to settle the debate. This line of defence sustains the viability of animal agency until proven otherwise.

Empirically found event regularities, such as apples falling to the ground, are commonly expressed as physical laws (here, the laws of gravity). These are used to characterise the workings of the Universe. A reductive worldview, in particular, would collaterally dismantle Steward's version of agency since her agents are supposed to possess properties which are not reducible to or supervene on basic physical properties. Steward claims many metaphysicians tend to show 'naturalistic sanity' with respect to basic physical properties. She argues that they often do this without outlining their concepts' relations to them whereas she abstracts from these properties altogether. This comes to the fore again later (§3.2.1).

Scientifically, it is arguably not possible to realise the falsity of the thesis of determinism on its own, reductive terms. We could only hope to spot higher-level causation indirectly in correlative sub-agential events. Any observable phenomena, then, do not necessarily pertain meaningfully to animal agency. Even if it is possible to settle the question about determinism, as I have discussed later (§3.2.2), there are significant experimental problems regarding accuracy and we might only be able to prove agency's *absence* when experiments are perfected and knowledge is whole. This collection of facts, by itself, delivers a major setback to the Epistemological Argument while granting hope to Steward.

There is another, more-epistemic concern with physical laws. Can empirical findings on the subject of agency be trusted? Physical laws possibly just represent abstract expressions of relations in nature that can only be tested indirectly (Popper 1959), whereby we revise our hypotheses (e.g. with new theories) and auxiliary hypotheses (e.g. and their conjoined claims) according to empirical findings. Take the acceleration of objects falling under gravity. This force is observable on Earth as an

increase in velocity of around 9.81 metres per second per second of a free-falling object near ground level, regardless of its mass. Who could reasonably deny this? But this Newtonian conception of gravity, which stems from ideas in 'classical physics', does not reveal the fundamental workings of the Universe, only facilitates somewhat accurate calculations. It is thought now, however, that Einstein's general relativity does. Meanwhile, countless galaxies across the Universe disobey our current laws of orbital mechanics since their 'arms' rotate at unexpectedly high speeds, which we speculatively attribute to unobservable, hidden mass called 'dark matter'. Additionally, to physics we have even introduced 'virtual particles', which allow us to account for many particle interactions, and 'imaginary time', which is routinely used in the languages of special relativity and quantum mechanics as a mathematical necessity.

Some scientists think it is sufficient to affirm an understanding of the Universe *in this way*. On imaginary time, Stephen Hawking wrote the following:

'One might think this means that imaginary numbers are just a mathematical game having nothing to do with the real world. From the viewpoint of positivist philosophy, however, one cannot determine what is real. All one can do is find which mathematical models describe the universe we live in. It turns out that a mathematical model involving imaginary time predicts not only effects we have already observed but also effects we have not been able to measure yet nevertheless believe in for other reasons. So what is real and what is imaginary? Is the distinction just in our minds?' (Hawking 2001, p. 59)

So, in harmony with Hawking, might constant scientific interrogation eventually lead to established facts? Undoubtedly, concepts such as imaginary time help us express what we observe in the world more accurately than we did prior to their inceptions. But even if some set of empirically found relations mounted a strong challenge to animal

agency in the present day, Steward's view might re-emerge in the future.[8] Yes, the predictive power of a new physical law might be great but, arguably, such epistemic revisions only spot more event regularities than before and not find permanent truths about the Universe's basic physical properties. A good scientist will accept new findings with an open mind, no matter how counterintuitive they seem. However, they should also realise that the increased predictive powers of expressed abstract relations do not necessarily equate to fundamental workings of the Universe that can be confidently extended to agency. In line with the Epistemological Argument and equipped with today's 'facts', *any* attack on animal agency is impotent.

Steward argues that her view is at odds with science's reductionism, putting herself under philosophical pressure to provide an alternative view of nature. If she cannot, her theory will lack sufficient reason to be intelligible (§3.2). So Steward appeals to a top-down order of causation. In particular, she cites the views of Nancy Cartwright (1999), who suggests an alternative to the usual reductive view of causality (Figure 1). Cartwright claims the Universe could be a 'dappled world' and encourages us to adopt a more-chaotic patchwork of explanation in its place, where our laws

'...do not take after the simple, elegant and abstract structure of a system of axioms and theorems...[They are] apportioned into disciplines, apparently arbitrarily grown up; governing different sets of properties at different levels of abstraction; pockets of great precision; large parcels of qualitative maxims resisting precise formulation; erratic overlaps; here and there, once in a while, corners that line up, but mostly

[8] Moreover, we can always doubt what empirical findings fundamentally mean *today* since they have a fundamental relationship with time. Intuitively, it seems absurd to defy the explanatory power of the scientific worldview like this. It would be akin to denying that the sky will be blue tomorrow, which we usually personally verify every day. But, even then, knowledge of the sky's blueness only pertains to what we have *deduced* about the world so far (as understood through our senses and Rayleigh scattering). Meanwhile, *inductive* proof will not help, for it requires its own induction, which is circular.

ragged edges; and always the cover of law just loosely attached to the jumbled world of material things. For all we know, most of what occurs in nature occurs by hap, subject to no law at all. What happens is more like an outcome of negotiation between domains than the logical consequence of a system of order. The dappled world is what, for the most part, comes naturally: regimented behaviour results from good engineering.' (Cartwright 1999, p. 1)

Given the disunity of this dappled world, disciplines such as psychology would be fully autonomous with respect to physics, whereby fundamental physical laws need not be applicable to our understanding of brain function. Hence Cartwright offers a non-reductive view of the Universe which opens a platform for a pluralistic notion of causation (§3.2.1). But why believe this view over reductionism? Cartwright, of course, could be wrong as well. Nonetheless, in citing Cartwright, Steward puts to us that we need not place faith in a reductive epistemic order of things through science to explain causal relations. As such, if the world cannot be understood reductively, we have even more reason to move on from the Epistemological Argument since only laws that pertain more-directly to agency, such as neuroscience, would be relevant.

Even then, results do not necessarily have power beyond the conditions in which they were found. Physical laws of causal relations have *capacities,* as Cartwright puts it. Laws are relativised to local test conditions (test populations, assumptions, etc.) and characterise particular sets of experimental observations. Whilst capacities are stable, laws are subject to change, depending on test conditions. These 'localised' laws only describe what regularly happens in our experience and do not represent the universally true, 'regular associations or singular causings that occur with regularity' (Cartwright 1999, p. 4) we are after.

Capacities represent the *potential* of the Universe to reveal stable event arrangements. If we postulate such a worldview, capacities would apply generally to all causal relations. Taking aspirin, for instance, carries the *capacity* to cure headaches. Increasing the concentration of carbon dioxide in the atmosphere carries the capacity to raise global temperatures. A law of relation we hypothesise following an experiment would only express a functional relation or change in probability between *local* causes and effects that do not carry to *all* test situations.

Only when laws are expressed with respect to their capacities can we provide a means to universally quantify causal relations as part of an untidy causal network. There is no reductive order to this system. Unlike abstract laws capacities are real and ontologically basic and come at all different levels. As they are not reducible to the causal laws they accompany, they cannot be induced in isolation. They *are* measured when we test causal claims and already expressed in the probability of an outcome from situation to situation. Naturally, however, laws need to be discovered afresh each time.

My grander point is that we do not owe trust to physical laws beyond the conditions in which they were derived; and *if* we are to eventually capture the structures of action behind causation of all kinds of movements in scientifically understood terms, we may need to look to capacities as well as physical laws for all test situations across the animal kingdom. The important point for Steward is that such a non-reductive order of things would not necessarily undermine animal agency like a reductive worldview would.

### 3.1.3. Science's fallibility

So where does this leave the Epistemological Argument? Our current scientific approach might never reveal universal determinism. The hope of scientists like Stephen Hawking is that the Universe is entirely governed by physical laws that *could* eventually be exposed with greater scientific knowledge and understanding. Even armed with this worldview, however, we can only *hope* to reveal the truth of determinism since, without a complete scientific understanding *now*, we have to beg the question in favour of determinism. Moreover, in contrast with any reductionism an opponent of Steward might rely on to contest her higher-level agency, an untidier piecemeal approach to scientific study might be more befitting of the patchwork we are probing.

It *could be true* that a theory of everything, dictated by something like quantum mechanics, underpins all physical movements in the Universe, even if we never uncover all its laws. Certainly, this would preclude self-movement in accordance with Steward's Agency Incompatibilism. However, this *could be false* too. The truth is that we simply do not know. Thus the Epistemological Argument, I agree with Steward, does not deal a fatal blow to animal agency, either now or prospectively, due to science's lack of superior command on the subject of determinism until convincingly demonstrated otherwise. As such, Steward's view, safely intact on an island of agnosticism, survives the attempted takedown.

In the meantime, there is much left to debate on Steward's notion of agency. While her top-down causation is a logically permitted candidate to explain animal agency, she might find herself question-begging if she cannot overcome the next hurdle.

### 3.2. Intelligibility

In overcoming the Epistemological Argument Steward liberates her theory from science's divine authority, so to speak, on the thesis of determinism. But now *she* must provide the necessary philosophical grit to make her theory intelligible.

By using 'intelligible' I mean to ask Steward why we should believe animal agency is real. That is, we require reason, ground, or cause to make her theory stick out above other accounts of freedom. This might epistemically promote her theory to become a fact of reality that can be developed further. I argue that, upon further inspection, this is where she encounters a snag. Specifically, I have put forward the following.

Steward's theory relies on substance causation in order to make sense of top-down causation. I argue that, by doing so, she comes up with conceptually clear metaphysical structures of action (§3.2.1). However, there are gaps in her theory. Furthermore, she reaches an empirically indemonstrable notion of agency since she abstracts agency away from the basic physical properties agents engage, leaving her notion of agency an unprovable concept (§3.2.2). I canvas a solution: can Steward adapt the Principle of Sufficient Reason (PSR) to 'bottom-out' these structures in terms of brute or unexplainable facts (§3.2.3)? I conclude that this would constitute begging the question in favour of her pluralistic notion of causation.

#### 3.2.1. Substance causation

Steward departs from the idea that agency can be explained in terms of conglomerations of sub-agential phenomena by cashing out something irreducible. Furthermore, she denies any supervenience, which can be defined as follows:

Supervenience: *Given two sets of properties, A and B, there cannot be an A-difference without a B-difference.*

The two sets of properties here are those of agency and basic physical properties. The question of how the properties of agency are conferred to animals if not from basic physical properties, then, still lingers, and Steward's view remains vulnerable to attack. However, Steward, of course, knows that animals exist in physical form since they comprise physical objects like eyes and fingers. But, she claims, these objects only confer properties to them in the sense that agents require these properties to self-move. Causation cannot be said to exist in those properties because, while they matter to agents, they do not *do the doing,* so to speak; rather, they, as universals, are merely *involved* in the events being brought about by agents and cannot be treated as sources of action.

Correspondingly, Steward looks to substance causation. 'Substance' generally pertains to an ontologically fundamental entity whose properties (e.g. colour, solubility, and magnetism) cannot exist without that entity. The bodies of animals, therefore, qualify as a substance which can possess relevant agential properties.

With her notion of substance causation, Steward appeals to a pluralistic causality, where causation is separated into at least three ontological categories: 'movers', 'matterers', 'makers-happen'. Movers are substances or collections of substances, matterers are spatiotemporal properties that causally relate to one another, and makers-happen are events that *trigger* other events such that the events are causally related. Take this sentence: 'The water splashed upwards because the alligator moved with aggression.' The alligator is the mover, the matterers are the relevant physical

properties (e.g. the water's viscosity and the momentum of the alligator's body), and the makers-happen are the particular events that prompt other events (e.g. the alligator's *shaking*, which led to *splashing*).[9] 'Causation' is an umbrella ontological concept that brings together these categories into a network of relations, where no category is given causal analysis.

While the animal agent, the mover, settles *which* movements to take it is pluralistic sets of relations—owing to movers, matterers, and makers-happen—that *causally engage* events. Cleverly, this tactic moves the nexus of causation away from agents. The semantics of 'agent causation' puts the agent at the metaphysical centre of causation. As such, an opponent can justifiably ask what makes the *agent* cause, where any explanation of causation from the agent requires its own explanation, thus initiating an infinite regress of explanation.[10] However, Steward is wise to this and posits that animal agents do not *cause*: the movements, themselves, as active events, do, (e.g. when an alligator triggers *how* it will give its bodily substance the 'kinetic oomph' to move and splash water). Steward successfully crafts an exit from the infinite regress. But Steward still does not explicate *how* these ontological categories are united in a pluralistic network of causation. She does recognise these gaps. However, with such an admission, she puts her theory's intelligibility at risk through its fundamental opaqueness. She must, then, fill these gaps with more than a passage led with the following:

[9] The giveaway of matterers is that they're usually connected with sentential expressions such as 'because'.
[10] For this reason Steward is wary of moving her position under the term 'agent causationism'.

'The main intellectual obligation of the causal pluralist is, I think, the explanation of what *unites* these ontologically various categories of cause. It is an obligation I cannot fully discharge in a book whose main concerns are elsewhere…' (Steward 2012, p. 211).

Steward must also explicate how agents, as substances with *non-physical* properties, trigger causal work over basic *physical* properties without being related to them. By disregarding reductionism and supervenience Steward has to appeal to an argument of absences, where causation is something that is hidden from basic physical properties. This puts her theory at risk of being unintelligible since we frequently make sense of the physical world we inhabit scientifically, notwithstanding my earlier comments on the scientific method's fallibility (§3.1).

To exemplify how relations to basic physical properties are not required Steward describes the causation of a whirlpool. Basic physical properties (molecules, interactions, etc.) still have an integral role in causation since they provide the required 'basal conditions' for the whirlpool's emergence. However, in this analogy the whirlpool, like an agent, is an abstract entity that is not reducible to or supervenient upon these properties: it is an entity whose higher-level organisation *coincides* with arrangements of them. Nothing is truly causal about its individual physical states or properties that relate to them. To claim otherwise is to beg the question in favour of determinism such that its individual states are enchained. The whirlpool as an entity and the forces that pertain to it, therefore, can only be understood abstractly.

Some philosophers, however, claim that nothing can act above its basic physical properties. Paul Humphreys (1989) posits that purely physical objects cannot be capable of causing by invoking John Stuart Mill's Methods of Agreement and Difference (Mill

1970).[11] To exemplify Humphrey's position consider the following sentence: 'The car demolished the wall.' In English language we give causal analysis to 'the car'. However, the car, despite being heavy, could not have caused the wall to be demolished without exhibiting significant momentum; thus Humphreys rejects giving causal analysis to the car because it requires a non-physical cause of its momentum. Steward thinks that this is absurd. Firstly, we do not *usually* infer negative conclusions (e.g. the car's causal analysis) from negative claims (e.g. the car's speed). But, more importantly than the semantics of causation, there is a pertinent methodological point to make (Steward 2011): proving relations between physical objects ought to be restricted to drawing empirical findings with respect to physical relations in nature. These cannot be used to service metaphysical ends in ontological inquiries when investigating the nature of agency which is abstract with respect to these relations. Whilst we can study relations between cars of certain momenta and particular walls, these objects are of the same ontological category. Any candidate relations they share, then, can only be discerned using Mill's deductive law or according to some laws *within* a general domain (e.g. physics), not those of metaphysical causation.

While Steward does not give causal analysis to physical objects, like agents, by themselves, she brings them together with matterers and makers-happen in a causal network that brings about actions. An agent, however, can only settle matters, according to this notion of causation, when causally sufficient conditions provide the

[11] In accordance with Mill's method causes can be identified when new regularities in events are observed after one variable in an experiment is changed. Mill, in one case, considers a bird that suffocates. The bird is moved from its cage into a container of carbonic acid prior to its death. The new physical properties of the bird's new spatiotemporal location and the gas are the only differences between the first state, a bird innocently sitting in a cage, and the second state, its suffocation. Therefore, it was caused to suffocate by its immersion in carbonic acid gas.

physical possibility. What are 'causally sufficient conditions'? In the case of the car and the wall they are things like physical laws (e.g. conservation of momentum) and conditionals (e.g. 'The wall will only be demolished if the car travels above 30 miles per hour'). These properties *matter* since they influence the probability of outcomes but they do not have causal efficacy by themselves; and only movements settled by agents—placing a foot to a pedal, turning the steering wheel rightward, changing the gear to '5'—begin the causal engagement of the car and the wall and its demolition. No single set of basic physical properties underlying the agent exhibits causal power nor do any overlying agential properties reduce to or supervene upon them: the capacity to settle is irreducibly abstract.

However, there is a complication. Is it possible for causation (a) within a substance ('immanent' causation) or (b) causation between entirely distinct substances ('transeunt' causation) to interfere with settling and, therefore, undermine causality which is strictly top-down, from settling to action? Let us say that (a) the driver's headache made their driving reckless and, further, that (b) they would not have been in the car had their mother not asked them to pick her up from the train station. Steward would claim that such events *do* interfere with what matters are settled but she would deny causal efficacy outside the agent's (driver's) substance. As with the cat and the moth earlier (§1.2.1), (a) and (b) are merely prompts or triggers that *constrain* what kinds of matters the agent can settle.

Given some charity, Steward paints a coherent picture of causation, despite some fundamental unknowns. But, now I ask, how can she claim, intelligibly, that animal agents actually exist without being able to demonstrate agency?

### 3.2.2. Demonstrability

Let me now add a serious, epistemic problem that puts into question how we can ever satisfyingly elucidate animal agency to render it an intelligible fact. While scientists attempt to provide intelligible descriptions of physical phenomena—sensible descriptions whose patterns of results can be empirically demonstrated repeatedly—Steward jettisons the scientific method for explaining animal agency. Not only is it not inevitable that the scientific method will ever demonstrate agency with them, empirical findings might only relate to local test conditions (§3.1). From a philosophical perspective, this encumbrance is frustrating: if the Universe's true metaphysical structures of action *are* empirically indemonstrable like this, knowledge with respect to agency is forever unobtainable, impeding epistemic advancements and casting the free-will debate to the permanent fires of speculation.

Steward claims in one short subsection of her book (Steward 2012, p. 223) that we already have *experiential* evidence of substance causation. She claims that an objection to this is one of fallacious Humean epistemology, according to which our ideas about causation come from expectation and not knowledge about the real world (Hume 2008). In this worldview animals have no role in causation as the Universe is characterised by series of events, unfolding deterministically or probabilistically, in a constant evolution of effects. 'Action', in this worldview, is simply a human construct alongside 'causation'. The creation of the Universe is its only ever cause until something intervenes (say, to destroy it). Steward strictly opposes such Humeanism, for it undermines substance causation altogether.

To counter subscribers to Humean epistemology Steward claims that they hide the very thing they are looking for. *Viz*, we interact with metaphysical structures of

action when we decide to pull a suitcase and succeed: we experience this relation. She also points to transeunt causation we can *observe* between lower-level substances, using a relation between mosquito eradication and incidences of malaria as an example. But, I contest, Steward's riposte here is scarce and, worse, flawed. Experience is just that: experience. Just because we believe we are causing such-and-such and just because we think we observe lower-level causation it does not mean that we actually do. At best, by treating causation as self-evident like this Steward tacitly concedes that the required metaphysical structures of action cannot provide a framework for demonstratable action outside *personal* experience. Unless she can allay these concerns in her project, Steward's view will be found epistemically stumped by her begging the question in favour of a pluralistic notion of causation.

One argument which makes the task even more difficult for Steward is considered by Randolph Clarke (2003). The argument is that anything purportedly held up as evidence for substance causation can be used as evidence for event causation. This is based on the premise that we only observe event-effects, not agent-effects. Steward does not account for differentiation between the two since she forsakes scientific probing. A barn owl might decide to fly towards a vole at time *T* in way *W* at place *P* because it intends to eat it. This makes logical sense as a set of movements that can be settled by the barn owl.[12] But, we must ask, how do we *demonstrate* that the owl brings about the decision to fly in the first place such that it initiates new causal chains of events to do work over physical properties (of its body and the vole's)? We do not see the own's influence on those *events*. Steward, of course, claims that the causal

[12] 'Fly' here is equivalent to '$\varphi$', which could be taken to mean 'Flap its wings and make use of gravity', or some other $\psi$. However we describe its movements, the owl must be assumed to have the same intention between $\varphi$-ing and $\psi$-ing when settling.

engagement of events is transcendent with respect to them. But this, unfortunately, renders her agency unobservable.

The fundamental issue is that Steward's argument of causation is one of *absences*, whereby causation is a network of connections, as described by her notion of substance causation. Animal agents bring about movements in their bodies but their bodies' properties are causally engaged by transcendent causality with respect to them. Jonathan Bennett (1988) holds an antithetic position on causation: that of 'concrete' facts located in spacetime and not abstracted away. In Bennett's view the transcendent 'oomph' behind movement is replaced by lower-level things like subatomic particles and aggregates of them, 'shoving' and 'forcing', which spatiotemporally *interact* according to facts of the physical world. This logic is applied all the way up to body parts, such as limbs, which immanently cause the body to move.

So either causality is concretely or it is abstractly held. Who is right? Can we prove *something*? Maybe we are wrong to think we can. Robert Northcott (2019) claims that free will (and, by association, animal agency) is not a scientifically testable hypothesis (e.g. through neuroscientific methods), for it is impossible to prove the causation of physical events without entailing prior physical events. As such, substance causation would not be directly verifiable, even in events. A correlation, say, between adrenaline levels and scatty decision-making, can be hypothesised and attested empirically but the nature of agents, in making the decisions over these sub-agential parts, cannot be revealed. 'Causation', Steward (2011) says, is too often conflated with 'explanation' like this, where correlated events are semantically confused with causal ones.

Inflaming this concern, Northcott posits that the only way to test free will is via proof of its absence, whereby one can only *falsify* the verification condition of freely willed neurological events. As such, we would need to possess a *complete* understanding of the brain's functions, which we may never reach. Additionally, the required accuracy to reach such a conclusion is likely unobtainable. Northcott presents many cases of shoddy accuracy even for testing the simplest human actions. By claiming that a theory of free will may never be directly proven Northcott compounds the idea that science is not necessarily in command to eliminate Steward's notion of agency (§3.1).

The corollary of all of this is that animal agency cannot be scientifically struck down by empirical findings until there is a complete understanding of how animals' movements are dictated by neural events. However, this does not help Steward as freedom becomes undiscoverable, much like the existence of God. Her arguments might remain starkly contingent forever in the sense they cannot proven to be correct or incorrect by scientific methods. If anything, then, Steward actually *loses* intelligibility where empirical demonstration is concerned and she needs to elucidate animal agency over other theories of freedom through other means.

### 3.2.3. Insufficient reason

To raise the intelligibility of animal agency we can look to PSR, which can generally be expressed as follows:

PSR: *For every fact,* F*, there must be a sufficient reason why* F *is the case*.

PSR in this formulation stipulates that her theory must have sufficient reason—a reason, ground, or cause—to be considered a fact. *F*, in this case, is the fact of animal agency, which Steward attempts to provide reasons for. However, since we cannot empirically demonstrate the truth of animal agency, direct scientific proof of *F* is out of the game (§3.2.2). Fundamentally, this seems to be because Steward's pluralistic notion of causality deploys an argument of absences, where causation is a

'…flexible umbrella concept under which we bring a wide diversity of ontologically various relations and relationships, unified only by their connections to our interest in the explanation, prediction, and control of phenomena.' (Steward 2012, p. 210)

In being an 'umbrella concept' she takes causation to be non-physical in origin such that it cannot be empirically demonstrated (nor can its relations be tested or even fully explained). Steward is then forced to rely on experience of causation to grant sufficient reason to it. But, at best, this restricts its corroboration to personal *experience* alone. Though she outlines a framework for metaphysical structures of action, why, substantively, should we believe hers are a fact of *reality*?

Perhaps PSR is too stringent altogether and we can ignore it. Animal agency could be a simple, *tout court* property of the Universe with no qualification. However, having no reason, ground, or cause is troublesome for Steward, for it relegates her theory to the realm of unintelligibility.

Alternatively, Steward could seek qualification of *F* in a contemporary version of PSR (Dasgupta 2016) if it takes the form of PSR*:

PSR*: *For every substantive fact* Y *there are some facts, the* X*s, such that (i) the* X*s ground* Y *and (ii) each one of the* X*s is autonomous.*

That is to say, according to PSR*, *Y* requires grounding by autonomous *X*s, where *X*s are essentially brute or unexplainable facts. *Y*, for example, could be the mental properties of hearing and *X*s could be the basic physical properties of acoustic sounds and ears and the brain which explains sound in terms of vibrations and its wavelengths and amplitudes and nerve impulses.

The scientific community accepts unintelligibility of autonomous facts like this, primarily in quantum mechanics, as we saw earlier (§3.1.2) with electron non-locality and imaginary time. Whilst these two concepts rest on theories for intelligible explanation, these theories rest on counterintuitive facts about the physical world (i.e. that electrons can be in more than one place at once and time can be imaginary). We bottom-out the explanations of many physical phenomena using the autonomous facts of physics yet we do not try to give sufficient reasons for them: they *just are*. After all, the origins of the Universe's own existence remain unexplained, owing to a scarcity of explicatory reasons or substantive evidence of a cause. Analogously, meanwhile, on one account of morality (Wielenberg 2009), there may be *sui generis* objective ethical facts that do not reduce to natural or supernatural facts. As such, moral claims can be made intelligibly with respect to these ethical facts.

Equally, then, Steward could claim that animal agency is a substantive fact that is apt for grounding by the autonomous facts of Steward's pluralistic notion of causation. Thus PSR* facilitates an intelligible explanation of animal agency, where the metaphysical structures of action—a network of top-down causation connecting

movers, matterers, and makers-happen—require no explanation. These structures circumvent the need for self-explanation or infinite regresses of explanation since they are brute or unexplainable facts about the Universe and how it came to be.

I think PSR* is Steward's best option. It provides the foundation to make a truth-claim on top of *something*. However, it requires significant charity: using PSR* would mean blankly accepting autonomous facts. Seemingly, we return to Square 1 since animal agency fundamentally remains unprovable. Steward has to gamble on the Universe possessing the right kind of fundamental properties to house her notion of causation. For all the knowledge we have about the Universe, every fact boils down to the same mystery of *how* things came to be. Therefore, Steward has to take a leap of faith since the truth of animal agency is *contingent* on facts about causality, which *could* be bottom-up in nature.

In summary, Steward argues for agents which are entities that are abstracted away from basic physical properties and are incorporated into a pluralistic notion of causation. I argue that this makes animal agency indemonstrable, a problem which is compounded by an unknown purportedly connecting her ontological categories, threatening unintelligibility. This epistemic fault at the fundamental level might not be hers—or, indeed, anyone's or anything's—but Steward *does* have to beg in favour of her notion of causation, which she overconfidently claims to be demonstratable from our experiences.

## 4. Where does this leave top-down causation?

Prior to writing this paper I embarked on an attempt to find a theory of freedom that was conducive to my wish for the debate to epistemically advance. Helen Steward's theory of animal agency was my choice. Having completed my investigation, I am not confident that my wishes have been fulfilled.

I first set out the great free-will debate before critically unpacking Steward's arguments, in which she claimed rudimentary freedom is available to even simple-minded creatures. Animal agents are centres of subjectivity and intent such that they can purposively settle to move their bodies at time *T* in way *W* at place *P* to initiate actions in face of open futures. Then I investigated how her theory deals with two objections.

On the first challenge, an opponent may level the accusation that Steward cannot claim, with integrity of argument, that animal agency is true whilst the thesis of determinism is an open epistemic possibility. However, I stand with Steward, who cogently argues her way out of it. Animal agency is safe on its island of agnosticism until her opponents explain *why* this open epistemic possibility is a problem. My view is bolstered by the idea that we have reasons to doubt the efficacy of the scientific method, especially of the reductive kind, in delivering the truth of the thesis of determinism, lending more credence to *philosophical* argument.

I then moved onto the second challenge, for which I asked Steward to provide sufficient reason for believing that its metaphysical structures of action are a fact of reality. I showed how she reaches a dead end when describing her ontological categories of causation, which is incomplete. Furthermore, since Steward discards scientific methods for this area of inquiry and deploys an argument of absences to delineate a

pluralistic causality which is transcendent in nature, I claimed that she abstracts animal agency away forever. I compared her argument to an argument for God's existence.

In my view, then, we may never come to ascertain the truth of Steward's idea of actions such that moral responsibility can be properly ascribed to human decision-making. This is worrying for the epistemically focused goals I initially held for this investigation. The only hope, I suggest, lies in the acceptance of brute or unexplainable facts. Such 'hope' requires the begging of Steward's metaphysical structures of action into existence.

## Bibliography

Anscombe, GEM. 1957. *Intention*. Oxford: Blackwell.

Bennett, J. 1988. *Events and Their Names*, Indianapolis: Hackett Publishers.

Broadie, S. 2013. Agency and Determinism in *A Metaphysics for Freedom*, *Inquiry*, vol. 56, no. 6, pp. 571–582.

Cartwright, N. 1994. *Nature's Capacities and Their Measurement*, Oxford: Clarendon Press.

— 1999. *The Dappled World*, Cambridge: Cambridge University Press.

Clarke, R. 2003. *Libertarian Accounts of Free Will*, New York: Oxford University Press.

Dasgupta, S. 2016. Metaphysical Rationalism, *Noûs*, vol. 50, no. 2, pp. 379–418.

Davidson, D. 1971. *Agent, Action, and Reason*, eds. R Binkley, R Bronaugh, and A Marras, Toronto: University of Toronto Press. Repr. in Davidson 1980, pp. 43–61.

—1980. *Essays on Actions and Events*, Oxford: Oxford University Press.

Dennett, D. 1971. Intentional systems, *Journal of Philosophy*, vol. 68, no. 4, pp. 87–106.

Frankfurt, H. 1969. Alternate Possibilities and Moral Responsibility, *The Journal of Philosophy*, vol. 66, no. 23, pp. 829–839.

Hawking, S. 2001. *The Universe in a Nutshell*, New York: Bantam Books.

Honderich, T. 1988. *Mind and Brain: A Theory of Determinism: Volume 1*, Oxford: Oxford University Press.

Hume, D. 2008 [1748]. *Enquiry Concerning Human Understanding*, ed. Jonathan Bennett, Copyright © Jonathan Bennett 2017.

Humphreys, P. 1989. *The Chances of Explanation*, Princeton: Princeton University Press.

Kane, R. 2002. Some Neglected Pathways in the Free Will Labyrinth, *The Oxford Handbook of Free Will*, Oxford: Oxford University Press.

Mill, JS. 1970 [1872]. *A System of Logic*, 8th edition, London: Longman.

Moore, GE. 1993. Moore's Paradox, *G. E. Moore: Selected Writings*, ed. T Baldwin, London: Routledge.

Northcott, R. 2019. Free Will is Not a Testable Hypothesis, *Erkenntnis*, vol. 84, no. 3, pp. 617–631.

O'Shaughnessy, B. 1980. *The Will: A Dual Aspect Theory: Volume 2,* Cambridge: Cambridge University Press.

Pereboom, D. 2013. Free Will Skepticism and Criminal Punishment, *The Future of Punishment*, ed. TD Nadelhoffer, Oxford: Oxford University Press.

Popper, KR. 1959. *The logic of scientific discovery*, New York: Routledge.

Steward, H. 2011. Agency, Properties and Causation, *Frontiers of Philosophy in China*, vol. 6, no.3, pp. 390–401.

—2012. *A Metaphysics for Freedom*, Oxford: Oxford University Press.

—2014. Precis of a metaphysics for freedom, *Res Philosophica*, vol. 91, no. 3, pp. 513–518.

Sartre, JP. 1958 [1943]. *Being and Nothingness*, trans. HE Barnes, New York: Washington Square Press.

Strawson, G. 1994. The Impossibility of Moral Responsibility, *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 75, no. 1/2, pp. 5–24.

Weatherford, R. 1991. *The Implications of Determinism*, London: Routledge.

Wielenberg, EJ. 2009. In Defense of Non-Natural, Non-Theistic Moral Realism, *Faith and Philosophy*, vol. 26, no. 1, pp. 23–41.

*The number of words in this paper (including the abstract but excluding other front matter and the bibliography) is 11,960.*