

# What Would Teleological Causation Be?<sup>1</sup>

John Hawthorne and Daniel Nolan

As is well known, Aristotelian natural philosophy, and many other systems of natural philosophy since, have relied heavily on teleology and teleological causation.

Somehow, the purpose or end of an object can be used to predict and explain what that object does: once you know that the end of an acorn is to become an oak, and a few things about what sorts of circumstances are conducive to the attainment of this end, you can predict a lot about the sprouting of the acorn and the subsequent behaviour of the piece of vegetation that results. Once you know that a rock seeks to move towards the centre of the Earth, you gain some insight into why it falls when released, and why it deforms the carpet or foot that it lands on. Once you know that the rabbit seeks to preserve itself, you can predict it will run from the fox. And so on.

There are at least three features of Aristotle's teleology, and more generally of an Aristotelian frame of mind about teleology, that may induce suspicion. One is that an end can serve as a "cause": as well as the sort of causation we all recognize, efficient causation, there are other forms, one of which is teleological causation.<sup>2</sup> However, this can look less odd if we think of causes as things that figure in "because" answers to "why" questions. Whether or not self-preservation, or the rabbit's continued existence, or something similar, causes the rabbit to run, the reply "because it seeks to continue in existence" certainly makes sense as an answer, or part of an answer, to a question about why it ran from the fox. (At present we are only claiming that it makes sense – we postpone the question of whether it is strictly speaking correct or particularly informative.)

---

<sup>1</sup> Thanks, to Frank Arntzenius, Cian Dorr, Hilary Greaves, and Hud Hudson for helpful discussion, to a reader for valuable and detailed comments. Thanks to John Carroll, Stuart Brock, and the audience at the 2005 Bellingham Summer Philosophy Conference for comments and critique. Thanks to The Ammonius Institute for research funding that helped John write this paper. Finally, thanks to Dean Zimmerman for providing the party that helped spark this paper.

<sup>2</sup> We do not wish to enter the debate about whether Aristotle's "four causes" are properly thought of as *causes*, or whether the four somethings-or-other are best understood as something else. That teleological "causation" was thought of as a kind of cause seems true of parts of the Aristotelian tradition, at least, whether or not it is a misunderstanding of Aristotle.

Another is the suggestion that things other than agents are influenced by teleology, and that objects can have these ends or purposes non-derivatively from the ends or purposes of agents.<sup>3</sup> That human beings or gods might have purposes that are not explained by the workings of efficient causation is not so mysterious – the idea that some of the things we do are not produced by any antecedent cause but are predictable or explained by purposes and ends we have is not unfamiliar. Nor is it an unfamiliar idea that artifacts can have ends or purposes derivatively.<sup>4</sup> A pen is for writing, because we make pens to have things to write with. A chair is for sitting on, rather than being for, e.g., burning or throwing, in part because its maker made it with the intention that people might sit on it. (And if we were theistically inclined, we might conjecture that some object or project has an end or purpose derivatively from the intentions of God. Some natural disaster was for the purpose of bringing people to their senses, for example.) But that a rock could be *for* moving towards the centre of the Earth, or that we could come across a piece of nature that, independently of any human or divine actions, was *for* providing food to humans, is far from how we typically think of things today.

Of course, the attribution of natural ends, ungiven by human or divine agency, is not entirely absent from our ordinary habits of thought. A tree's roots are partly to draw moisture and nutrients from the soil, and partly to anchor the tree, and maybe there are other ends they serve as well. But we need not think that the tree is some sort of proto-agent, sending its roots out intentionally, and most of us at least understand the thought that trees might be around without the deliberate creation by a natural or supernatural agent. One's heart is for pumping blood, and even though it's around as a result of action by agents (e.g. parents), we don't think that they designed it or gave it its purpose.

---

<sup>3</sup> See also Hegel: 'One ought not to conceive the end under the form it assumes in consciousness, -- that is to say, under the form of a representation' (*Logic*, Section 104) quoted in Paul Janet, *Final Causes*, Scribner's, 1892, p. 346.

<sup>4</sup> Though Bayle thought there to be something deeply misleading about this way of thinking: 'But if a faculty without consciousness and reason merely because it is created by an intelligent being, becomes fit to accomplish works that require intelligence, is it not as if it were said that, of two men equally blind, the one does not know his way, the other knows it because he has been created by a father with eyes? If you are blind, it matters little whether you were born of a blind or seeing father, for in both cases you always need to be guided by the advice and the hand of another.' (Quoted by Janet, p. 359.)

There are various stories we can tell about this “natural teleology” – we might think histories of selection explain why it is correct to attribute purposes to tree-roots, or we might think the hypothesis of deliberate divine creation explains why everything that has a “natural” purpose or end, has that end. Of course, we need not keep these explanations in the back of our minds when we agree that tree roots are to draw up moisture and not to spy on rabbits. But when contemporary philosophers and biologists tell stories about this natural teleology they tend to proceed as if there is a different underlying explanation: superficial teleology gives way to an underlying reality that is not fundamentally teleological at all. This is so even in the case of mental activity. Teleology gives way to mental representations that play efficient causal roles (which in turn may enjoy yet deeper explanations that proceed via categories that are not mentalistic at all).

This brings us to a third, and perhaps most noteworthy, feature of the Aristotelian framework, one that is the most troubling of those that we will consider. Teleological explanation can be fundamental: it need not itself be true in virtue of some underlying efficient causal facts.<sup>5</sup> We are happy enough with stories about purposes and ends when there is some deeper efficient causal story to be told in the background. When we agree that a heat seeking missile flew north because it was seeking heat (e.g. a plane’s engine or exhaust), we assume that an explanation of the inner workings of the missile in terms of infra-red radiation hitting detectors prompting electrons to flow around and electric motors moving internal workings around is also available to explain why the missile shifted direction and headed north. If we were told that no such explanation was available, and that Boeing had cracked the secret of how to make missiles just plain seek heat without such workings, we would be bewildered.

Contemporary orthodoxy does not, then, take the idea of fundamental teleology very seriously. What we would like to do in this paper is to explore some questions about what it would be to take some cases of teleological causation to be fundamental: that is, to take some cases of teleological causation to not require explanation in some other terms (e.g. in terms of an underlying network of efficient causal laws). Despite the importance of the Aristotelian tradition in this, we are not particularly concerned with Aristotle exegesis. We think the concept of final

---

<sup>5</sup> We shall clarify the relevant notion of ‘fundamental teleology’ in due course.

causation answers to something in our pre-theoretic conception of the world (or at least is a natural development from that conception of the world), and also to ideas deployed in many theories besides Aristotle's own. In asking questions about what best sense could be made of this concept, we take ourselves to be asking a somewhat different question to the question of how Aristotle deployed it, or even what Aristotle's theories would look like after a little cleaning up. We should emphasize, in particular, that we have no intention of clarifying teleology by recourse to the kind of matter/substantial form metaphysic within which teleology was historically framed -- where, roughly speaking, the end of a thing is determined by a substantial form, that in turn does not supervene on the material properties of a thing.<sup>6</sup> It is no accident that fundamental teleology was abandoned at the same time that matter/form metaphysics gave way to the "catholick affections of matter".<sup>7</sup> But it remains far from clear that fundamental teleology needs the kind of metaphysical underpinnings from which Boyle, Descartes, Gassendi, and others recoiled.

### ***Why Care about Fundamental Teleology?***

Some of our readers will be asking why they should care. Isn't fundamental teleology part of a superseded, pre-scientific muddle about how the world works? Indeed, haven't reflective minds known this for quite a long time? Centuries ago, Francis Bacon remarked: "Inquiry into final causes is sterile, and, like a virgin consecrated to God, produces nothing."<sup>8</sup> Perhaps there are some whose intellectual interests invite or compel an interest in fundamental teleology. For example, those who feel obliged to think that Thomas Aquinas is basically right about philosophy seem to face an especial burden to make sense of fundamental teleology. But as for the rest of us,

---

<sup>6</sup> And thus the substantial form serves as the formal cause of the final cause. It is one thing to attribute an end to a thing, quite another to say that it is fixed by a substantial form that does not supervene on the ordinary physical properties of a thing.

<sup>7</sup> See Robert Boyle, 'About the Excellency and Grounds of the Mechanical Hypothesis,' (1674), reprinted in Thomas Birch, ed., *The Works of the Honourable Robert Boyle*, Volume IV (1772).

<sup>8</sup> *De Augmentis Scientiarum* Bk iii Ch. 5, quoted in Woodfield, *Teleology*, Cambridge University Press, 1976, p.3.

aren't questions about final causation about as interesting as questions about phlogiston or angels dancing on the heads of pins?

We think there are some good reasons to be interested.<sup>9</sup> The first reason that one might be interested is that many people think that it was a matter for a posteriori discovery that there was no fundamental teleology. This was something that science revealed to us about the world, not something to be settled by a priori cogitation. If it is an a posteriori question, and one that we settled in the negative through scientific investigation, then what part of our evidence, exactly, bore on that question? If it is an eminently a posteriori matter, shouldn't we think that some possible courses of experience would count in its favour? If so, what would they be? If fundamental teleology is not to be dismissed a priori, then what is our good reason for dismissing it? Those interested in how general scientific questions are born upon by our empirical evidence should be interested in how our evidence impacted on such a high-level issue as whether there was primitive teleology.

On the other hand, the issue of whether there is fundamental teleology is interesting if it is not, after all, an a posteriori matter to be settled by general scientific inquiry. For the question of whether there is fundamental teleology looks like a general hypothesis about the nature of the world that is not obviously inconsistent (and perhaps not inconsistent at all) – and if general hypotheses of this sort are to be settled, if at all, by a priori means, then a priori cogitation has more of a role in the sciences than many are at first inclined to think. It would be interesting in understanding the history of science, at least, if the biologists of ancient Alexandria were being held up by having made an important a priori mistake, rather than the sort of mistake to be discovered and corrected only by a posteriori methods.

There are more subtle alternatives here that might be considered. Perhaps someone might think that whether there is fundamental teleology is not an a priori matter, but that while a posteriori evidence could count against it, no a posteriori evidence could count in its favour. Perhaps one could even think that it is discoverable a priori that no course of evidence could favour fundamental teleology. If one of these subtle options is true, that might cast some interesting light on the relation of theories to evidence or the role of a priori epistemology. One way or

---

<sup>9</sup> We are interested in phlogiston and angels on the heads of pins, too, but we recognize some of our readers will have less broad interests.

another, final causation seems unlike more ordinary cases of commitments discarded in the light of scientific investigation. The kinds of problems we have in working out what is evidence for or against the existence of phlogiston or Vulcan or smallpox-causing demons seem different, in that it is relatively straightforward to gesture at what sort of evidence could have counted for or against these hypotheses, but the relation of the hypothesis of final causation to evidence is much more of a philosophical puzzle.

So those prepared to dismiss fundamental teleology as having been ground under the wheels of advancing science may be interested, given their position, in the articulation of how advancing science gave us reason to reject it, or alternatively to explain the a priori faculty that would have allowed us to dismiss it in advance of the a posteriori evidence. Others may be interested in the question of final causation because they are less sure that it ought to be consigned to the flames. This may be particularly plausible if one is a non-reductionist about scientific inquiry, and one thinks that very different sorts of investigations have their own fundamental theoretical commitments: even if fundamental teleology has no place in contemporary physics and chemistry, it is much less obvious, if one is a non-reductionist, that it has no place in contemporary zoology or ecology or psychology or sociology. The question of fundamental teleology is arguably to be settled on an inquiry by inquiry basis: and we may want illumination about what final causation amounts to in order to decide whether there is any in areas where it might be thought to be a live option.

We have been talking as if fundamental teleology is a dead option in physics and chemistry, but even this might be challenged.<sup>10</sup> There was certainly a revolt against it and other scholastic commitments at the beginning of the modern period, but we are not entirely sure how late live teleological alternatives existed even in physics. Consider anti-atomistic construals of the laws of thermodynamics, which remained respectable into the twentieth century. Suppose we construed closed thermodynamic systems as aiming at thermodynamic equilibria, and that the equilibria that the theory postulates are the ends towards which the system is tending,

---

<sup>10</sup> See, for example, von Weizsacker, *The World View of Physics*, London: Routledge and Kegan Paul, 1952.

in a way that is not to be further explained in other physical terms. Prima facie, that theory looks rather teleological.

Perhaps some contemporary trends in physics are as well served by teleological interpretations as any other. Systems of laws dominated by “principles of least action”, for example, seem to lend themselves to descriptions of systems primarily in terms of what they tend towards, or what they “aim at”, everything else being equal (or ideally taking everything else into account).<sup>11</sup> Some would want to gloss this talk as being at best metaphorical, but in advance of some reason to reject the postulation of fundamental teleology, it is not clear why we should have to.<sup>12</sup>

Finally, some of our readers may be motivated to explore fundamental teleology for more esoteric reasons. They may still be attracted to Aristotelian or Thomist conceptions of the natural world. Or they may want to consider teleology an option in theology: however a simple divine psychology operates, it is presumably not in virtue of a host of discrete underlying efficient causal mechanisms, so maybe the purpose and end-seeking that we offer underlying accounts of human neuropsychology are brute in divine psychology. Or perhaps they are attracted to certain idealist or post-idealist metaphysical pictures: perhaps they believe that there is a *Weltanschung* that has its evolution explained and predicted teleologically; or they take the parts of our world to always and already be infused with the purposes for which humans use them, and that this “ready-to-handedness” is not to be further explained in terms of the interaction of human psychology and a non-teleological world, perhaps because the latter are abstractions from the former which is taken as fundamental. One of us at least feels no need for these hypotheses, but we invite our readers who do have these commitments in joining with us in thinking about what primitive teleology would be.

---

<sup>11</sup> This continues a thread that began at least as early as Leibniz (see, for example, his contributions to *Acta Eruditorum* in 1691 and 1697 on the catenary curve and the brachistochrone problem), who provides a fascinating case study for those interested in teleological thinking. See especially Jeffrey K. McDonough, ‘Leibniz on Internal Teleology and the Laws of Optics,’ ms.

<sup>12</sup> A perhaps more interesting argument for certain least-action principles being non-teleological would be that they are time reversible.

### *Why Suppose Our Question Has an Answer?*

Those readers who have persisted might have been convinced that our question is interesting. But there seem to be some reasons to think that it will not have an answer. Perhaps final causation is metaphysically impossible. Perhaps “final causation” is a natural kind expression that fails to correspond to an actual natural kind, so even if it is conceptually coherent it still lacks enough meaning to be employed usefully. Perhaps it could be used to argue that questions about final causation are, after all, and perhaps even a posteriori, semantically defective in a way that precludes answers to our question from being anything but indeterminate. Or perhaps final causation is possible, but there is no fact of the matter about what it would be like, were there to be some. At the extreme, one might even think that facts about final causation have no necessary connections with any other facts (or very few), so anything could be teleologically produced by any end whatever.

The discussion that follows is designed, in part, to answer these abstract challenges. What we hope to do is to describe nomic structures that first, are readily intelligible, and are thus immune to a charge of obscurantism; that second, are (plausibly) metaphysically possible; and that third, are good enough deservers of the name “teleological laws”. That is to say that, all things considered, if a world was to enjoy fundamental nomic structures of the sort that we are about to describe then it would enjoy enough of the structure implicated by teleological thinking as to make true an ascription of fundamental teleology to it. There may yet be better possible candidates for primitive teleology than the ones we will describe: but we have done enough for this paper if we show that the notion is coherent and improve our understanding of how it could be applied. We are not going to say that our model is the only way that a world can have teleological laws – quite the opposite – but only that it is one good enough way.

### *Some Ground Clearing*

To begin it is worth noting that some standard philosophical decision points about efficient causation carry over to teleological causation. In the efficient case, there are those who see their task as that of providing an account of efficient causal laws, reckoning particular instances of causation to be manifestations of some governing

law or other. And there are those who are singularist in orientation, allowing for the intelligibility of efficient causal connections between events that do not proceed via any overarching laws. So in the case of teleological causation, we might on the one hand take our task to be one of understanding the nature of teleological laws, assuming that any particular instance of teleological causation will implement some such law; and on the other, we might allow that something might do some activity for some end without this having anything to do with some covering law. At the extreme, one might even allow for duplicate worlds – qualitatively and nomically, one of which contains an object that does A for end E, another of which contains a duplicate object that does A for a different end (or for no end at all). The topic of fundamental teleology is difficult enough without attempting to accommodate rampant singularism at the fundamental level. We shall thus set ourselves the task of getting a handle on what would reasonably count as a fundamental teleological law.

Another standard contrast within the space of positions on efficient causation is between reductionism and non-reductionism. There are those who believe that the causal facts supervene on a ground floor that can be perspicuously described without recourse to causal and nomic concepts. A paradigm case in point here is David Lewis, who believed that the facts about which causal laws obtain at a world and what causes what at a world supervene on facts about the spatiotemporal distribution of qualities at that world.<sup>13</sup> Since our focus is nomic, it is especially useful to recall his view about what makes it true that a set of laws is fundamental, namely, that it is the set of generalizations that best combines simplicity and informativeness. At a rough first pass, to gauge simplicity and informativeness, we examine a formulation of the laws in a language in which the most natural properties correspond to lexically simple predicates, and less natural properties are expressed by complexes. Simplicity is then gauged by the length of the total expression of the laws, and informativeness is gauged by some natural measure on how much is ruled out by the law. We are by no means committed to this sort of reductionism about laws of nature. But it would certainly aid the cause of fundamental teleological laws were they presented in such a way as to be compatible with reductionism – thus requiring no strange causal-telic

---

<sup>13</sup> See especially *Philosophical Papers Volume II* (Oxford University Press, 1986) and *Papers in Metaphysics and Epistemology* (Cambridge University Press, 1999), where this picture plays a frequent and systematic role.

primitive relations at the metaphysical ground floor. We will thus present a conception of teleological laws that is *compatible* with reductionism. Better still, we will present an account of fundamental teleological laws that is compatible with Lewis' gloss on fundamental laws: there are possible worlds for which fundamental teleology provides the best combination of simplicity and informativeness. Teleological causation and laws could be perfectly coherent, and perhaps even possibilities, even if they were irreducible. The advantage of providing a description compatible with reductionism about laws is that we can appeal to a broader audience.<sup>14</sup>

### ***Teleology and Backward Causation***

There are some initial thoughts we want to dispose of about what counts as fundamental teleology. One initial thought is that we have teleology exactly when current states are systematically nomically correlated with future states. Aristotle's acorn, which seeks to become an oak tree, has its behaviour predicted and explained, it seems, with reference to its future state: specify what a mature oak tree is like, and tell me that the acorn will change so as to move closer to that state, and you're on the way to enabling me to predict what the acorn and young oak will do in a variety of situations – the roots will sprout down and the leaves will sprout up, the sapling will get taller and wider each year, and so on. When we treat a heat-seeking missile teleologically, we keep in mind that its "aim" is to impact a strong heat source in its vicinity – the twists and turns of a heat-seeking missile following a fighter-jet engaged in evasive action can be predicted when we think of the missile as "trying" to hit the plane's engines.

---

<sup>14</sup> Even having provided an account of teleological laws, there remain tricky issues about how to deal with the causal relation between particular events. One might think that an earlier event causes a later event only if that sequence is 'covered' by the efficient causal laws. But such a thought risks relying on the currently unfashionable deductive-nomological account of event causation. Whatever account we do opt for, we should remember that in the teleological tradition, man thought that an earlier event couldn't cause a later event without there being final causation, since what makes an event suited to cause what it does is something final. Thus Thomas Aquinas says in *De Principiis Naturae*, 4.22 "Whence the end is the cause of the causality of the efficient cause". On that view, then, teleology explains efficient causation between events. We shall not pursue these issues further here.

There are several reasons why we should not just take teleology to be a matter of nomological correlation with a future state. One is that if the laws are suitably deterministic (so that the state of the world at a time determines both a unique future and a unique past), a specification of the future state of a system together with the laws will determine the past state of a system – and we would not want Newtonian mechanics to count as teleological simply because it is deterministic.

There is another kind of conceivable prediction and explanation which relies on specifications of a future state to predict and explain past states, and which we think is worthwhile to not confuse with teleology. Many people think that it is at least conceptually possible that there be backwards causation (backwards efficient causation, that is). Perhaps the best known thought experiments about backwards causation are time travel ones: my packing a sandwich now before stepping into my time machine explains the presence of a sandwich covered in plastic wrap in the twelfth century, and if you know how the insides of my time machine are now, you can better predict and explain the goings-on inside my time machine when it arrives in the twelfth century. Despite this, in these sorts of cases we are not inclined to think that the way the sandwich is in medieval times is explained by some fundamental purpose it has to then be a certain way in the early twenty-first century.

Consider a sequence of events like the following: we have a black box travelling at a constant velocity through space, surrounded by a miscellany of objects. At some point, some objects behind the box have particles leave them on a trajectory that will impact with the box (so they are emitted at angles that will strike where the box will be, not where it is at the time of emission). Shortly afterwards, more objects have particles that leave them behind the box, together with some objects well ahead of the box's path. All of these particles are moving with velocities and trajectories so that they will "impact" the box at the same time. As the time of intersection of the various paths gets closer, objects nearer and nearer the box have particles leave them. Finally, all the particles touch the surface of the box at the same time and disappear, and it turns out that their momentums all cancel out (and let us suppose the box's mass-energy changes at that point to become the sum of its previous mass-energy plus those of the particles). If that is what we knew of the behaviour of this system, what sorts of laws might we conjecture governed the movements and causal reactions of those particles, the objects they came from, and the box?

We should give some credence to this being a massive coincidence. We could give some credence to the hypothesis that this was all a massive set-up, and that some past causal process “primed” the situation to produce a coincidence in the future, of many particles striking the box at once from all directions. We could, if we thought it made sense, ascribe primitive “purposes” to these particles, or perhaps the objects they came from, to seek the box at the time of intersection. Or, finally, we could describe it as a case of backwards causation – the “intersection” event was actually an emission event, when the box released many particles travelling backwards in time, and as a causal result of this, the particles were absorbed by the objects in the box’s environment some time before the emission. In both of the latter cases, the state of the box at “intersection time” would in an important way explain and predict (or retrodict) the movements of the particles. But the style of understanding this sequence of events as involving emission of time-travelling particles seems very different from attributing primitive teleology.

We think we have some grasp on when it would make sense to interpret these events as being a result of backwards causation (though to make the choice might require more information than we are given just by the description above). For example, if we found particles like that in our world, and they conformed to the equations of relativity theory, given our confidence that those laws are near-enough right in describing the velocities and mass-energies of particles, we would have reason to believe that they were moving backwards in time rather than forwards at faster than the speed of light. If we further had a reasonable conjecture about how such particles were produced, and we found the box was set up as predicted while the objects that the particles were first observed next to were a physical miscellany which were not plausible emitters of these exotic particles, then the case for backwards causation of the particles by the box would be very strong.

This example suggests that when we have a course of events where past states seem to be nomologically dependent on future states and not vice versa, we should not rush to a teleological explanation of the future states. If we discover that the behaviour of acorns now is predicted and explained by the flourishing of mature oaks later, one possibility we need to rule out is that those oak-states-of-affairs are causing the acorn-sprouting states of affairs through backwards efficient causation. When the missile flies towards the heat-source, one option to discount (though normally we discount it without a second thought) is that exploding jet engines spit out heat-

repelled backwards-in-time-travelling missiles, which explains why there are missiles in missile cradles some time before the relevant sorts of engine exploding.

It is worth pausing to characterize the difference between backward causation and final causation. One point of divergence has to do with the stepwise character of backward causation. In the standard cases, when some future state  $s_1$  backward causes some past state  $s_2$ , there is some chain running from  $s_1$  to  $s_2$  such that proximate members of the chain stand in backward causal relations to each other. The Aristotelian picture of final causation is not like this. The development of a tree at a given time is explained by the mature state, but not by the state right after that time.

Of course one might not insist that backwards causation absolutely must proceed in the stepwise fashion just alluded to. In the case of past to future efficient causation, some of us wish to allow for action at a temporal distance, so that some event at  $t$  causes some event an hour after  $t$  without there be a connecting causal chain. Similarly, then, one might wish to allow backwards causation at a temporal distance. Why not treat final causation in this way?

This last suggestion still overlooks one absolutely crucial feature of teleology, namely that final causation is supposed to be compatible with the end not actually being reached, so there need not be any later state for the earlier state to be nomologically correlated with at all.<sup>15</sup> The acorn sprouts and turns into a sapling even if it gets cut down before it becomes a mature oak. The rock tumbles to the bottom of the hill, even if it never reaches the centre of the Earth. The agent saves money to buy a house, even if a bank crash bankrupts him before he has enough. An object can act because of its aim, or its function, even when it never reaches its aim or its function cannot be successfully carried out. Whatever final causation would be, it does not seem that it could be a matter of something being brought about by a future event, unless non-existent future events were as able to bring things about as well as existent

---

<sup>15</sup> Of course some may allow efficient causal laws to connect properties  $F$  and  $G$  in worlds where  $F$  is instantiated and  $G$  is not. But the point is that in the efficient causal case, the law explains by way of connecting two events (the cause and the effect), whereas in the final case, application of the law will not require a pair of existing events as relata.

ones. To the extent that we find causation by non-existents absurd, we should hunt for another way of getting a grip on final causation.<sup>16</sup>

(A further reason to distinguish final causation from nomological dependence on future states of affairs is the coherence of the notion of backwards final causation, where the transitions of a system are explained in terms of an aim or function concerning its state in the past. Backwards final causation has never been discussed, so far as we know. We shall focus, with orthodoxy, on forwards final causation, though it will be obvious how to extend the discussion to backwards final causation.)

### *Aristotelian Rocks and Telic Pyramids*

The Aristotelian tells us that rocks fall in order to be at the centre of the Earth. Even in this simple case, there are a number of implicit assumptions that are worth bringing out, ones that will help us a great deal in articulating a serviceable version of final causal laws.

First, as we try to adopt an Aristotelian perspective on the case, we naturally suppose there to be a number of jointly indeterministic non-final constraints upon the rock. We suppose that it is not possible for the rock to move discontinuously. The final causal law doesn't explain that. Nor is it clear that the final causal component of nomic space will explain the limitations of speed on the rock's descent. And so on. On the other hand, the non-final constraints on the rock cannot themselves determine a unique path. Otherwise there would be no work for final causal laws to do! In general, when final causation explains the path of a system, there will be non-final constraints upon the path.

Second – and this is crucial – there are possible cases in which we wish to say that a rock acts in order to reach the earth despite the fact that its actual motion was far from the best way to reach the earth. Indeed, its motion may have made it radically less likely to reach the earth. Suppose the rock begins by being dropped just above the

---

<sup>16</sup> One way to keep the idea that the final states always do causing in cases of final causation would be to employ Aristotle's distinction between existence in actuality and existence in potentiality. If we said that there was already a mature oak existing in potentiality as soon as the acorn came into existence (or a mature-oak-state-of-affairs existing in potentiality) that mature oak (or oak-state-of-affairs) could be the final cause we need. To the extent one wishes to reject a kinds-of-existence doctrine, one will find final causation so described as unappealing.

lip of a suspended open bucket. Suppose the “constraints”, whatever they are, do not determine whether the rock will go down or sideways. Now if the rock first went sideways and then went down it would do a far better job at reaching the earth than if it headed straight into the bucket. How, then, is one entitled to say that the rock went down in order to reach the earth? If the law says “Do what it takes to get to the earth,” then the rock will violate the law. Meanwhile, if the constraints require the rock to fall, then it will be the constraints and not the teleology that explains its falling.

It is of no use at this point for the Aristotelian to tell us that the rock “thought” that going down was the best way of reaching the earth – and that the laws say that rocks will do what they think is the best way of reaching the earth. This response would confirm the already significant suspicion that Aristotelian teleology is bound up with an illicit projection of mental states on to unthinking things.

The challenge can be answered. Doing so brings to the fore the kind of ideological structure that will be appropriate to rigorous development of the final causal picture. First, given a system, there will be a state that is thought of as the end state. Second, one will deploy a natural metric for calibrating the system’s distance from the end state. (Let us be clear, distance is not being thought of as something that in general correlates with physical distance from some object or place. For this reason, the rock example can be somewhat distracting.) Now once such a distance metric is in place, we can define an abstract notion of end-velocity in terms of (abstract) distance and time. Whatever the relevant notion of distance the coordinate notion of end-velocity will be such that end-velocity (which can be positive or negative) over an interval is a matter of the change in distance from the end point divided by duration of the interval, and the end-velocity at a time is the derivative of distance over time.<sup>17</sup>

Return to the rock hovering over the bucket that is released. We could imagine a law saying:

L1 If a rock has one or more paths to the end state compatible with the constraints, it will take a path (that is compatible with the constraints and) which gets

---

<sup>17</sup> Note that since there is a privileged end state, there are only two directions relevant to the vectorial component of end-velocity: towards and away from.

it to the end state at least as quickly (in time) as any other path compatible with the constraints.

But note that this gloss on the law – one that makes no reference to the notion of end-velocity and hence doesn't depend on any suitable notion of distance – does not accord at all with the Aristotelian picture of the teleological behaviour of rocks (or, relatedly, with the actual behaviour of rocks). The rock goes down and not sideways, but the law recommends otherwise. Moreover, in a situation where the constraints preclude the rock reaching the ground (suppose it to be in surrounded on all sides by a suspended sphere), it will still be a fit subject for teleological explanation.<sup>18</sup> But L1 will be silent in a case where there is no available path to the end state that is compatible with the constraints.

Imagine, instead, the following law (to be made more precise later):

L2 At any time  $t$ , a rock will follow a continuation  $c_1$  of its path in such a way for every other continuation  $c_2$  compatible with the constraints, there is a period after  $t$  such that the rock has greater end velocity on  $c_1$  than on  $c_2$  during that period.<sup>19</sup>

If one is looking for metaphors, L2 invites the image that rocks act blindly towards their ends, while L1 suggests a picture according to which rocks act with foresight. As the Aristotelian thinks about it, teleological explanation as applied to unthinking matter is apt to be of the blind sort and so naturally regimented by L2.

We in no way mean to suggest, of course, that the most suitable system of fundamental laws for actual rocks will take the form of L2. But might there not be a

---

<sup>18</sup> Note that, similarly, Jonathan Bennett's account of teleological laws in *Linguistic Behaviour* (Cambridge University Press, 1976) does not make room for this kind of explanation. His 'basic theory' of teleological laws says "any R thing will do whatever is required for and sufficient for its becoming or remaining G" (p. 39). A law of this sort cannot explain the rock's behaviour in this case. Meanwhile, his refined theory of teleology involves a kind of proto-belief – 'registration' – which will be out of place for rocks. Of course, since Bennett is not trying to provide an account that could underwrite, say, fundamental teleology for rocks, these considerations may not spell trouble for his own project.

<sup>19</sup> Note that we do not mean to be arguing here that the only laws that deserve being called 'teleological' are ones that take the shape of L1 or L2.

world for which certain of the fundamental laws take the form of L2? Imagine a world of the following sort.<sup>20</sup> Now and then energy spawns clusters of distinctive particles – call them p-particles – that are distributed in all sorts of wild and wonderful patterns at the time of inception. The particles share a simple intrinsic structure. In general, we notice that the clusters of particles move in such a way that they form pyramidal structures, a fact that obviously cannot be explained in terms of the intrinsic features of the individual particles. Barring some violent external influence, clusters of particles form pyramids. And barring some violent external disruption, such structures stay in place once formed.

We investigate closer. We notice a fairly natural way of calibrating how close a cluster of particles is from being organized pyramidally, and that a correlative notion of end-velocity can be defined. We then notice that relative to such a calibration (and a plausible account of the other constraints on the system of particles), something very precise can be said about the behaviour of the system of particles: at any given time, the system acts so as to maximize end-velocity. We thus hypothesize that the following law obtains:

L2\*: For any system  $s$  of  $p$ -particles  $s$  time  $t$ ,  $s$  will follow a continuation  $c_1$  of its path in such a way for every other continuation  $c_2$  compatible with the constraints, there is a period after  $t$  such that the rock has greater end velocity on  $c_1$  than on  $c_2$  during that period.

A few obvious points can be made in connection with the example. First, it is clear enough as an epistemological point that it would be very natural for the people in our imagined world to regard L2\* as both fundamental and teleological. Second, if we accept the Lewisian story about what makes a fundamental law true of a world, we can very well imagine that some such law would figure – from a God’s eye point of view – as the best combination of simplicity and informativeness and thus count as a fundamental law. Third, even leaving to one side the Lewisian story, there seems to

---

<sup>20</sup> Another interesting kind of case to consider, suggested by Frank Arntzenius, is a time-reversed version of big bang worlds, where everything ultimately collapses to a singularity in systematic ways that are elegantly captured by teleological generalizations that take low entropy as the telos. (Of course, those who wish to reduce the direction of time to entropy flow might dispute the example.)

be no good a priori reason why such a law could not be fundamental. Of course it may be that as a matter of deep a posteriori (or inscrutable) necessity, a law of this sort is impossible. But if this is so, it is a modal discovery that has not yet been made.

One final observation: L2\*, as it applies to individual particles, is non-local, but this should not be confused with its teleological character. The non-local character consists in the fact that the behaviour of a particle at a time is not determined by its local environment but rather by the configuration of the other particles in the system. The teleological character of the law, meanwhile, consists in the fact that the behaviour of the system is to be explained in terms of getting closer to a privileged state (that may or may not be actualized). Non-locality certainly does not require teleology. And teleology does not require non-locality. (Imagine that point particles at some possible world had a very complicated set of intrinsic properties, and that the evolution of the internal life of a point particle was explained in terms of an L2 style law that depended upon some calibration of distance between possible intrinsic states of a point particle and some privileged state....)

### *A More Formal Presentation*<sup>21</sup>

Rigorous presentations of physical theories often proceed in terms of a configuration space for physical systems, whereby the possible configurations of a system are given as points in an n-dimensional space. Possible trajectories for the system (let us assume for now a classical notion of time) can be given as functions from times to points in the configuration space. The ideas presented above can, in certain cases, be formulated quite naturally in terms of some such formalism.<sup>22</sup>

Let  $\_$  be the configuration space of the system.

Let T be the set of times.

Let a path be a function f from T to  $\_$

<sup>21</sup> Thanks to Frank Arntzenius and Hilary Greaves here.

<sup>22</sup> This formalism is naturally applicable in a case where the relevant fundamental teleological law says that the world-system has an end, as opposed to teleological laws that posit ends for certain subsystems (i.e. where the development of a subsystem turns on how the entire system most quickly approach its end, rather than one accords each subsystem its own end.) It nevertheless serves a nice illustration of how certain kinds of teleological laws could be articulated in a rigorous way.

Let  $F_{\text{ant}}$  be the set of paths  $f$  that are compatible with the constraints.

Let  $C$  be a designated point in the configuration space. (Of course, it may be more natural in some cases to treat  $C$  as a region and not as a point. Think back to the pyramids – there is more than one point in the state space where a given cluster of particles forms a perfect pyramid. We shall ignore this complication in what follows.)

Let  $V_{tt'}(f)$  be the average end-velocity of path  $f$  between  $t$  and  $t'$ .

Let  $m(x)$  be a function from configurations  $x \in \_$  to the real numbers  $\mathbb{R}$ , where  $m(x)$  gives the distance of configuration  $x$  from configuration  $C$ .

An L1 type teleological law can now be represented as follows:

$$F_{\text{Tel 1}} = \{f \in F_{\text{ant}} : (\forall t \in T)(\sim \exists f' \in F_{\text{ant}})(\forall t' \geq t) f(t') = f'(t') \wedge (\exists t' > t)(f(t') = C) \wedge (\sim \exists t' \geq t)(f(t') = C)\}^{23}$$

(Hereafter we shall call laws of this type ‘quickest path laws’.)

Laws of an L2 type can be expressed by the following type of constraint on paths:

$$F_{\text{Tel 2}} = \{f \in F_{\text{ant}} : (\sim \exists f' \in F_{\text{ant}})(\exists t \in T) \forall t' \geq t) f(t') = f'(t') \wedge \exists t'' > t)(\forall t''' > t'' > t) \supset V_{tt'''}(f) \geq V_{tt'''}(f')\}$$

(Hereafter we shall call these ‘end-velocity laws’.)

We are prepared to consider both sorts of laws as teleological laws, though our primary interest is in the L2 type of laws.<sup>24</sup>

## Objections and Replies

---

<sup>23</sup> Here and in what follows further qualifications would be in order if one were to allow that in certain cases, there is, for every path, a quicker path compatible with the constraints. (One may, for example, write the law in such a way that it dictates nothing were this situation to arise.)

<sup>24</sup> For what it is worth our subjective probability of a fundamental law at the actual world being an end-velocity law is also somewhat higher than that for a quickest path law.

Objection One: Even if quickest path laws or end velocity laws obtained and were fundamental, that wouldn't make it true to say that there was final causation or fundamental teleology in the world in question, because the laws do not *say* that things act 'for' the sake of getting to C (the point relative to which the relevant notion of velocity is defined) or that have C as their end, or that they act with the purpose of getting to C. One can't simply *stipulate* that in such a case C is the end and that things act for the sake of C.

Reply: Quite often, when there is a disputed term or concept, the cogency of which is in question, one who wishes to defend the coherence of the relevant concept proceeds by describing situations using concepts that are not under dispute and then argues that the situation so described captures enough of the semantic intentions underlying the original concept as to make it reasonable to claim that were such a scenario obtain, the concept in question would apply. Abstractly conceived, the strategy is an altogether familiar one. The defender of the cogency of the concepts of say, human choice or external colours, describes the world in terms that are acceptable to all parties and then argues that such a world would fit the relevant facts about use well enough to render correct various predications of freedom and colour. Such an advocate might, of course, concede various idiosyncrasies or natural tendencies in our conceptions of colour or freedom that could not be made good upon. Perhaps we habitually think that if a surface is yellow, then so is every part of the surface (which runs into conflict with the fact that atoms aren't yellow). Perhaps we habitually think that if one has a choice about x and y necessitates x then one has a choice about y. But a failure to fit every such twist and turn in our conceptions of colour and freedom is not yet to render such predications false. Such a move space ought to be absolutely familiar to philosophical readers.

Suppose one thought that specifying a fundamental end-velocity law, for example, was not sufficient for teleological causation. The challenge would be to say what more was needed. Such a law would at least be evidence that there is teleology, and we maintain that the law by itself would be enough, though presumably our opponent would think more is needed. One kind of opponent we might face is a teleological mystic who takes the view that teleological facts and laws were such that they could not be capturable in any other terms: unless a possible scenario is specifically described in teleological language, one has not done enough to ensure that

there is anything teleological about it. By such lights, to deprive oneself of the concept of ‘acting for’ is *ipso facto* to become blind to teleological facts altogether. Such a primitivist might insist that there are two possible worlds governed by the same end-velocity laws, one containing teleology and one lacking it. She would not be sympathetic to our claim to have described a situation with teleology.

While it is difficult to say anything decisive against this form of primitivism, we are not sympathetic to it: it is one thing to allow that teleology be fundamental, quite another to insist that a teleological understanding of the world is so foundational and primitive that the facts about the world that make it correct cannot be described in any other terms. One puzzle for this sort of primitivism, for example, is to explain why lawlike behaviour of possible objects like our “telic pyramids” suggests teleology so strongly if it has no conceptual connection with it. We are more open-minded about the challenge that allows a teleological set-up can be captured in other terms, but that it requires more than we have proposed. Of course, while we are open-minded, we are yet to be convinced that anything more is needed.

Objection Two: If one is sneaky about one’s choice of a metric, can’t one always reformulate a law in terms of a teleological law that explains the world in terms of some abstract end velocity law? For example, can’t one select an arbitrary point in a Newtonian state space and then cook up a gerrymandered distance function  $N$  and a coordinate end-velocity law  $L$ , such that things satisfy  $L$  iff they satisfy the laws of Newtonian mechanics?

Reply: Minor point: there will be some principled challenges to translating some sets of laws into teleological form. In particular, if a set of laws is time reversible (so that if a history is permitted by the laws, then so is the time-reversed version of the history), then there will be a problem in principle in coming up with a translation scheme, since teleological laws of the sort that we have been describing are not time reversible. Major point: The problem is essentially no different from Nelson Goodman’s riddle of induction.<sup>25</sup> We know that, in general, a generalization that is framed in terms of “projectible” predicates will be analytically equivalent to a generalization that is framed in terms of gruesome, gerrymandered, predicates. But we

---

<sup>25</sup> See Fact, Fiction and Forecast, Harvard University Press, 1955.

do not standardly conclude that the both generalizations are equally good candidates for expressing a law of nature.<sup>26</sup>

Even those of us that do not approve of David Lewis' reductionism think that there is something right about the idea that the laws will be expressed in terms of predicates that express very natural properties. A similar lesson applies here. Not all distance metrics relating arbitrary points to a selected point in a state space will be equally natural: some will be deeply gruesome, others not. Generalizations that take the form of quickest path or end-velocity laws but which deploy highly non-natural distance metrics will not be very good candidates for expressing laws of nature. Meanwhile, it will certainly not be true that any system of efficient causal laws can be translated into a set of teleological laws that deploys fairly natural distance relations.

Of course, we are aware that some readers will be altogether wary of the Lewisian hierarchy of naturalness, and will urge that while not all properties and relations attract our attention equally, none are objectively more real, or natural, or haloed than others. It is clear what such readers ought to conclude: the question of whether the world is teleological is theory dependent. From this neo-positivist perspective, it makes no sense to ask whether the world is teleological or not, only whether some notation for expressing the world's regularities is teleological or not. We are not inclined to such a view. Those that are will be forced to concede that the choice not to describe the world using laws couched in the teleological forms we have described is a choice that is not forced upon us by the world but by some suitable mixture of biology, fashion, and convention.

Objection Three: This account of teleological laws does not capture Aristotelian teleology because it makes no mention of the good. After all, Aristotle tells us that we explain what happens in terms of what is for what, he imagines that such explanations will take the form 'because better thus – better not simply, but in relation to the reality

---

<sup>26</sup> For those people that individuate propositions in a coarse way, so that sentences true at the same set of worlds automatically express the same proposition, the point will have to be put differently: a gruesome generalization will express a law of nature only if there is way of expressing the same generalization using projectable/natural predicates.

of the thing concerned'.<sup>27</sup> For Aristotle then, reality is governed by the overarching principle that each thing tends towards that which is good for it. The present account makes no mention of this.

Reply: It is obviously right that the good plays an important role in Aristotelian teleological theory. Nevertheless, we think clarity is best served by proceeding in the way we have. For we think it is useful to separate the question whether any fundamental laws are end-velocity or quickest path laws from the question of whether such laws are instantiated by normative target points and distance metrics that correspond to some normative better/worse ranking. Let us agree, for example, that in order to come reasonably close to the Aristotelian vision in respect of teleology, it would not be enough that some end-velocity laws obtain. For the Aristotelian would no doubt hope that each relevant C would be the state that is best for the system, what constitutes flourishing for the system. Furthermore, the Aristotelian would hope that for any system (or at least any system that deserves the name 'substance' – let us not dwell on this subtlety here), there is a law that tells us that when other things are equal, the system heads towards its flourishing point. Moreover, in such a situation, the maximal blend of simplicity and informativeness would be achieved by a general law that said that any system will tend towards its flourishing point (where each flourishing point bears a natural similarity to any other with respect to the abstract – but far from gruesome – property for being what is good for the system that is at that point).

One reason that clarity is best served by the proposed separation of issues is that while some people will object to the idea that end-velocity or quickest path laws might be fundamental, others will confine their concerns to the hypothesis that flourishing points and normative better/worse relations might figure in such laws.

As we see it, it is eminently natural to regard at least end-velocity laws as teleological even if they are detached from notions of flourishing and even if, on that score, they do not vindicate certain important aspects of the Aristotelian tradition. In particular, it is natural to do so on account of deep structural analogies between

---

<sup>27</sup> Physics Book II, Chapter 7 in Ackrill, *A New Aristotle Reader* (Princeton University Press, 1987), p. 106. See also the way that Aristotle ties the good to 'that for the sake of which things are done' in the Nichomachean Ethics and elsewhere.

explanations that proceed via end-velocity laws and paradigmatically Aristotelian explanations.

Let us say that a world is Robustly Aristotelian iff a very large range of its inhabitants are governed by end-velocity laws in which the designated end state for a thing corresponds to what is best for the thing in question and where the relevant distance metric corresponds to what is better or worse for the system. We have indicated that we see no objection in principle to the possibility of worlds governed by end-velocity laws. What though of the suggestion that a world might be Robustly Aristotelian? There are two obvious kinds of obstacles to such a suggestion. The first is scepticism about the very possibility of the normative properties required for such laws – being good for a system and so forth. It will hardly be obvious to everyone that Aristotelian flourishing properties – which apply to non-biological systems and that hold of systems independently of the intentions of minded users and creators – are even possible. Second, even if such properties do exist, one might think them insufficiently natural to count as suitable bases for generalizations that deserve the name ‘natural law’. These, indeed, seem to be the fundamental objections. Once the relevant domain of natural properties is conceded, it will at least be rather difficult to see why a Robustly Aristotelian world is impossible (at least on a priori grounds). Once again we can take the Lewisian conception as an initial springboard. At least by Lewisian standards, it will be very hard to argue that Robustly Aristotelian worlds are impossible (once the above concession is made), and this is because it will be extremely difficult to argue for the metaphysical impossibility of a world in which suitably normative end-velocity laws are the best compromise of informativeness and simplicity.

### **Why Is There Something Rather than Nothing?**

We end on a lighter (some will think heavier) note. Turning to one of the ‘big’, time-honoured, questions in metaphysics, one might reasonably wonder whether teleological laws could explain why there is something rather than nothing. We incline towards an affirmative answer. Consider a world in which the following obtain: (i) The “constraints” in that world do not prevent something following nothing. (ii) The teleological laws entail that if a world is such that there is nothing at any time, then reality gets closer to C (the end given by the laws) by having not

nothing after that time than if reality stays at nothing. (Note that this requires an end-velocity law that applies to reality itself rather than to some subsystem within it.)

With this in place it is quite clear that it is nomologically impossible that reality always contain nothing. Since we are inclined to think (i) and (ii) are possible, we are inclined to think that at least at some worlds, there is a teleological reason why there is something rather than nothing. Moreover, if (a) time stretches back infinitely, (b) the constraints never require something becoming nothing, and (c) a move from a situation  $s$  that is not nothing to nothing is always a move further from  $C$  than a move from  $s$  to a situation that is not nothing, then it will be nomologically impossible that there is ever nothing.

Such explanations have obvious limitations. If the laws are contingent, then the explanation will not tell one that there *has to be* something rather than nothing. And if one wishes to liberalize ‘something’ to include time itself, the explanation will no longer be satisfying. In that case, if a teleological explanation can proceed at all, it will have to be embedded in an abstract state space in which not every point is one in which time even exists, but to which the by now familiar machinery of constraints and teleological laws can apply. We postpone such flirtations with the limits of understanding to another occasion.