Hugh Desmond, Philippe Huneman

The Ontology of Organismic Agency: A Kantian Approach

Abstract: Biologists explain organisms' behavior not only as having been programmed by genes and shaped by natural selection, but also as the result of an organism's agency: the capacity to react to environmental changes in goal-driven ways. The use of such 'agential explanations' reopens old questions about how justified it is to ascribe agency to entities like bacteria or plants that obviously lack rationality and even a nervous system. Is organismic agency genuinely 'real' or is it just a useful fiction? In this paper we focus on two questions: whether agential explanations are to be interpreted ontically, and whether they can be reduced to non-agential explanations (thereby dispensing with agency). The Kantian approach we identify interprets agential explanations non-ontically, yet holds agency to be indispensable. Attributing agency to organisms is not to be taken literally in the way we attribute physical properties such as mass or acceleration, but nor is it a mere heuristic or predictive tool. Rather, it is an inevitable consequence of our own rational capacity: as long as we are rational agents ourselves, we cannot avoid seeing agency in organisms.

Introduction

Stags lock antlers to gain access to mates. Arctic poppies rotate and track the sun in order to maximize solar exposure. Bacteria swim up a sucrose gradient in order to get better access to the source of sucrose. When biologists explain organisms' behavior by referring to their goals in this way, then they are using what we in this paper will call *agency explanations*. Such explanations make sense of organisms' behavior *as if* they were agents with goals.

Despite its philosophical pedigree (going back in some form to Aristotle), the problem of organismic agency was neglected in much of twentieth century philosophy of biology and mainstream evolutionary theory, which was dominated by the *artifact approach* to organisms. The organism was understood to be a collection of functional traits, designed by natural selection in much the same way Paley's watch was designed by an intentional creator. Thus, each pattern of apparently purposive behavior was understood to be a functional trait that is pur-

Hugh Desmond, University of Antwerp and KU Leuven
Philippe Huneman, CNRS / Université Paris I Sorbonne

https://doi.org/10.1515/9783110604665-003

posive in name only ('teleonomy', cf. Pittendrigh 1958; Ernst Mayr 1961). In philosophy of biology this view was enshrined by the 'selected effects' account of function, where all biological functions can be explained by a process of natural selection (Wright 1973, 1976; Millikan 1984; Neander 1991).

In recent decades a more robust approach to organismic agency, which we will call the *agential approach*, has become increasingly influential. Organisms are agents with goals and purposes that interact with their environments, and their behavior can only be understood with reference to the goals of organisms as wholes rather than as mere collections of parts (e.g., genes or traits). This focus on whole organisms (following Bateson 2005) is linked with several ongoing developments in evolutionary biology, most notably the so-called Extended Synthesis (e.g. Müller 2017).

This motivates taking a new look at the long-standing question whether or not organismic agency is 'real', in the same way that the wings of a bird are, or the claws of a bear. After all, agency ascriptions to organisms have long been suspected of being mere metaphors and fictions of the human mind – an anthropomorphic projection even – rather than an accurate description of the mind-independent world. We call this question the question of whether to adopt an *ontic view* of agential explanations (cf. Salmon 1989). In an ontic view of agential explanation, agential explanations explain because they refer to an element of the ontology of the world (i.e., agency) which is responsible for the explanandum (i.e., organismic behavior) – just in the same way that causal-mechanical explanations explain because they single out the actual mechanism that causes the explanandum phenomenon (Craver 2014).

Our approach in this paper will be to overlay this question with a distinct but closely related one: whether agential explanations ultimately can be reduced to non-agential explanations (a worry raised in e.g., Lewens 2007). For instance, once one asks the question why organisms have such-and-such purposes and not others in the first place, the agential approach rapidly becomes inadequate. Why do stags want access to mates in the first place? The most plausible explanation would seem to involve a selection explanation, along the lines of 'those stags who tended to not engage in sexual competition did not get to transmit their genes to the next generation'. Thus, the genesis of organismic purposes is explained through a process of natural selection. However, does this also imply that purposeful behavior can be explained without reference to organismic purpose or agency, without loss of explanatory power? We call this question the question of *explanatory dispensability*. Agency is (explanatorily) dispensable if and only if an agential explanation can be replaced by an explanation void of any reference to organismic agency or purposes, without any loss of explanatory power.

Even though there are four possible combinations of answers to these two questions, most contemporary thinking about agential explanation focuses on two. The first is the non-ontic (epistemic) view of agential explanation where agency is dispensable. Its main representative today is what we call the 'Neo-Fisherian option', which holds that agency is invoked in explaining behavior for purely heuristic reasons – in particular, as shorthand for other types of explanation (especially selectionist explanation). This option has been especially widely adopted in behavioral ecology where, following Grafen's 'maximizing agent analogy' (Grafen 1984), organisms' behavior is analyzed as maximizing inclusive fitness (Grafen 2006).[1] The Neo-Fisherian option can be traced back to Fisher's fundamental theorem of natural selection, which states that, under the influence of natural selection, populations of organisms have a tendency to increase fitness (equal to the population's genetic variance in fitness: cf. Fisher 1930, chapter 2). In this way, agential explanations could be adequately replaced by explanations that do not refer to organismic agency and purposes (but only to natural selection), and organismic agency therefore is not a mind-independent causal power in the way, for instance, natural selection is assumed to be.

The second option, decidedly less mainstream but increasingly defended, combines indispensability with an ontic view of agential explanation. We call this the 'Neo-Aristotelian option', since it expands fundamental ontology to include organismic purposes. Different versions of this option have been developed in recent years: most prominently by Walsh (Walsh 2012, 2015), but Moreno and Mossio's analysis of biological autonomy also follows the Neo-Aristotelian option (Moreno and Mossio 2015), as does Varela's notion of autopoiesis (Varela 1979).

In this paper we seek to identify a third option[2] which we call the 'Kantian option' regarding agential explanations: (1) the concept of organismic agency is indispensable to scientific explanation and (2) agential explanations are to be conceived non-ontically[3]. In particular, viewing organisms as agents with pur-

1 Inclusive fitness is a fitness measure that includes the (expectation of the) number of kin offspring (so a sterile individual could have a high inclusive fitness if its relatives had many offspring), mitigated by the degree of genetic relatedness between relatives; for this latter reason, the maximizing analogy forms a bridge between behavioral ecology and population genetics.
2 The fourth option – where agency is explanatorily dispensable and yet considered robustly real – is possible but does not strike us as particularly compelling. After all, if a concept is dispensable, Ockham's razor directs us to discard it from our ontology.
3 To what extent 'non-ontic' should be interpreted as 'epistemic' sensu Salmon (1989) is a rather complicated question which we discuss at the end of section 5.

poses is a "demand of reason": it is necessary given our rational nature. This means that attributing agency is not a consequence of our limited computational capacity, or of our contingent evolved nature that causes us to detect agency falsely (cf. the so-called 'agency detection' cognitive modules: Atran 2002; Barrett 2000). Yet at the same time, it is a mistake to believe that agency is a natural regularity or causal process, belonging to the 'furniture' of the world in the same sense as physical processes. In this way we will suggest how one can obtain the robust explanatory indispensability desired by (some) Neo-Aristotelians (i.e., agency is not just a heuristic) without the ontological price that Neo-Fisherians would be loath to pay.

The paper is structured as follows: in the first section we give a broad introduction to organisms and the major streams in biological thought, written for non-specialists (i.e., philosophers outside the philosophy of biology). In the second section we define with more precision what an agential explanation is and contrast it with functional explanations. In the third we discuss various attempts to replace agential explanations with non-agential explanations, and argue that – despite widespread hopes – once one looks at the details, one cannot but conclude that attempts to make agency dispensable, even today, remain aspirational rather than clearly successful. In the fourth section we discuss Kant's original approach to teleology in the natural world and show how it can be the basis for our Kantian approach to agency. In the final section we show how the Kantian approach entails viewing agential explanations as a 'demand of reason'.

1 Artifacts and Agents

Much of mainstream twentieth-century evolutionary biology operated within the framework of what is called 'the Modern Synthesis', a term coined by Julian Huxley (Huxley [1942] 1974). The Modern Synthesis was forged in the 1930s and 1940s by Ronald Fisher, Sewall Wright, Theodosius Dobzhansky, and John Haldane, among others, and is often described as the synthesis of Mendelian genetics and Darwin's theory of natural selection. It was very much focused on how allele (different versions of the same gene) frequencies change over time in response to evolutionary forces, such as natural selection, mutation, drift, or migration.

Organisms were essentially analyzed as epiphenomena arising from changes in underlying allele frequencies. In the words of Huxley ([1942] 1974), they were viewed as "bundles of adaptations" where each adaptive trait was shaped by natural selection in response to environmental demands – just as artifacts are

designed and put together, piece by piece, by an artisan. Such a view of organisms has never been unanimously accepted, even among the major architects of the Modern Synthesis (cf. Mayr 1982; Simpson 1944, 1953), but the view has nonetheless been the dominant one, and has been popularized in the work of Richard Dawkins (Dawkins 1976). Dawkins introduced a dichotomy between replicators (alleles) and interactors (organisms), with the consequence that organisms are mere tools in a never-ending arms race between genes, with genes the genuine actors in evolutionary history. Even apparently goal-directed organismic behaviors, such as beavers building dams, are expressions ('extended phenotypes') of the underlying genotype (Dawkins 1982). In sum, while it may *seem* that an organism undertakes behavior to further its own goals (e. g., secure food, fend off predators, etc.), it does so actually for the benefit of the genes, which get to replicate when the organism does well. In this way, the theoretical resources of the Modern Synthesis were used to support a philosophical view of agency as dispensable and fictional.

The metaphor of Paley's watch, which dominated in the early days of the Modern Synthesis,[4] was supplemented after the 1960s with analogies borrowed from computer science. Organismic behavior was often described as *programmed,* starting with influential papers by Mayr (1961) and Jacob and Monod (1961):

> The purposive action of an individual, *insofar as it is based on the properties of its genetic code*, therefore is no more nor less purposive than the actions of a computer that has been programmed to respond appropriately to various inputs. (Mayr 1961, 1504, our emphasis)

So even if the behavior of an organism may seem goal-directed, that is only because its genetic code has been 'programmed' by natural selection to direct the organism to react in certain ways to certain inputs, and in other ways to other inputs. Organisms are no more goal-directed than computers are.

Despite the metaphors of 'design' and 'program,' it is important to note that even biologists operating squarely within the Modern Synthesis were well aware of the limits of the metaphors. In the quote above, Mayr qualified the programming analogy with "insofar as it is based on the properties of its genetic code." Mayr is not claiming that individual organisms behave exactly like pre-programmed computers, only that some aspects of their behavior are determined by environmental inputs in the way that a computer program responds to user inputs. Similarly, in *The Extended Phenotype*, Dawkins devotes a whole chapter to debunking the view that genes determine all aspects of organismic behavior, a view he calls the 'myth' of genetic determinism.

---

4 Cf. Lewens 2005 for an in-depth discussion of the artifact metaphor.

The limitations of the artifact metaphor are built into one of the very foundations of the Modern Synthesis: the analysis of phenotypic variance as proposed by Fisher (Fisher 1919). This analysis states that, in general, only a part of the variation of phenotypes in a population is explained by a corresponding variation in genotype. The rest is variation in environment (impacting how the organism develops), or variation in how genotype and environment correlate (cf. e.g. Hamilton 2009).

Thus, no practicing biologist holds that organismic behavior (or phenotype) is entirely determined by a genetic program[5], for the very simple reason that the *environment* is the second element that goes into determining phenotype.

The role of the environment points to limitations in speaking about the adaptive 'design' of organisms. A genotype may be designed for a particular type of environment, i.e., there may be a particular 'normal' environment in which the bulk of the selection for that genotype occurred. In that normal environment, the genotype develops into an adaptive phenotype. However, in reality, environments are highly heterogeneous, so in a population of identical genotypes, only a fraction will develop in the 'normal' environment. Other environmental inputs – inputs that differ from the normal environment – cause the organism to diverge from its 'designed' phenotype. In this way, while theoretical resources in the Modern Synthesis lend some support to the artifact metaphor, the same resources point also to the metaphor's limitations.

Moreover, the role of environment in organismic behavior (and phenotype more generally) also provides a direct motivation for the agential approach. To see this in more detail, consider the phenomenon of phenotypic plasticity. A trait is 'plastic' (in the context of quantitative genetics[6]) when the underlying genotype can develop into different phenotypes solely due to environmental variation. The degree of plasticity of a trait is represented by the term V_E in the equation above.[7] Plasticity, defined in this way, is an incredibly basic phenomenon: it simply refers to how different environments cause genotypes to develop into different phenotypes. At its most basic, it can refer to phenomena that are the result of physical or chemical (rather than properly biological) processes, such as the stunting in the growth of a plant in response to poor nutrition. There are few if

5 Whether organismic behavior can be entirely explained by natural selection is a more difficult question, since the environment also can be influenced by natural selection through niche construction. We discuss this in Section 3.

6 There is also cell plasticity, referring to the multiple dispositions of a totipotent cell in developmental theory. This is not relevant here.

7 The term describes how different genotypes are correlated with different degrees of plasticity; or in other words, how different genotypes react differently to environmental novelty.

any organisms that lack some form of phenotypic plasticity in some of their traits.

The phenomenon of plasticity was not considered to be of any special significance until the work of Bradshaw (Bradshaw 1965); before him, the phenotypic variation due to environmental perturbation was often viewed as noise. Bradshaw showed that plasticity in a trait can be adaptive to heterogeneous environments. If an organism can vary a trait in response to changes in its environment so as to be able to adopt a more adaptive phenotype, such an organism can be at a selective advantage in variable environments, compared to an organism without that capacity. In particular, Bradshaw distinguished between four types of environmental heterogeneity where plasticity can be adaptive[8] (Bradshaw 1965, 21): (1) when the environment changes on a time-scale that is equal to or shorter than generation time; (2) when the environment varies over very short spatial scales; (3) when the magnitude of environmental variation is very large; (4) when it is beneficial to maintain a stable phenotype in a population while maintaining genetic diversity.

Adaptive scenario (4) shows how maintaining stable phenotypes in the face of environmental change is also a form of plasticity. When it becomes inscribed into developmental pathways, it has been termed 'canalization' (Waddington 1940); moreover, plastic maintenance of stable phenotypes is hypothesized to precede genetic accommodation where the phenotype is produced by genetically determined developmental pathways (West-Eberhard 1989). Finally, some degree of canalization in organismic traits is nearly ubiquitous, since thermodynamic fluctuations in the molecular bases of genes would be detrimental and then counter-selected if they were significantly affecting the development of phenotypes.

Adaptive scenarios (1)-(3) refer to organismic behavior that is often thought of as (apparently) agential. For instance, in response to chemical cues emitted by sea slugs, bryozoans will develop spines to defend themselves (Godfrey-Smith 1996). Such forms of plasticity open up parallels with cognition, and not surprisingly, theorists and philosophers concerned with the evolution of cognition often take the evolution of adaptive phenotypic plasticity to be a model (van Duijn, Keijzer, and Franken 2006; Lyon 2017; Calvo Garzón and Keijzer 2011; Sterelny 2000; Caporael, Griesemer, and Wimsatt 2013; Godfrey-Smith 1996). Organisms exhibit a whole range of cognitive, or at least apparently cog-

8 See also Nicoglou (2015) for the history of Bradshaw's study, and Desmond (2018) for a more detailed discussion of the role of temporal and spatial scale.

nitive[9], behaviors: they sense changes in the environment, are able to process this information and select a response from a repertoire of responses. Far from being a late-stage development in evolutionary history, we see these types of behaviors in bacteria, which can undertake evasive action upon detecting predators (Pérez et al. 2016), or swim to a food source upon detecting sucrose gradients (Auletta 2013).

In sum, the impact of the environment on phenotype shows – via the phenomenon of phenotypic plasticity – how it is not entirely adequate to view organisms simply as artifacts. Moreover, it motivates a definition of organismic agency as the capacity to respond to changes in the environment in such a way as to further organismic purposes. Organismic agency thus understood is a much broader concept than the agency traditionally ascribed to human, rational subjects, which is typically characterized by means of some mental state, like an intention (cf. Schlosser 2015). The approach to organismic agency in the biological sciences, by contrast, blackboxes whatever cognitive processing may or may not be going on. In this sense organismic agency is best understood as an ecological property (cf. Walsh 2015), namely, a property of the interaction between organism and environment. We will now discuss agency and agential explanations in more systematic detail.

2 Agential Explanations

Definition of Agency

For the purposes of this paper, we will operate with the following minimal working definition of agency:

A system is an agent if and only if (1) it possesses a certain purpose P, where P is a particular state of the system, (2) it maximizes the realization of P in response to environmental change, and (3) the system itself is a cause of the realization of P.

While we view this definition as being continuous with established work in this area (cf. Moreno and Mossio 2015, 92–93), it may be helpful to explain the various elements involved in the definition. Condition (1) models the purpose of a system as a particular state. For organisms, purposes may refer to developmental states, physiological states, or behavioral states. Condition (2) specifies that

9 The application of the term cognition, as well as other terms such as communication or memory, to organisms such as bacteria remains a controversial point. See discussion in Lyon 2015.

goal-directedness is to be interpreted as a maximization or optimization. This equation of purposefulness with some type of optimization is common across the sciences. Finally, condition (3) is intended to exclude clear non-agent systems, even where a process of maximization is occurring, such as the marble rolling down into the middle of the bowl (minimizing gravitational potential energy). Here the marble is not considered a cause of its own maximization behavior. The same is true of more complex physical systems, such as Bénard convection cells, which are patterns of heat flow that appear spontaneously when the temperature gradient is large enough. Such structures may maintain their organization even in the face of perturbation in their environment, such as movement of the container walls (Manneville 2006); nonetheless, they are widely considered not to be agents (Moreno and Mossio 2015). By contrast, an organism that modifies its phenotype in order to be more adaptive to a new environment is considered to be a cause of the modification of its own phenotype.

While 'self-causation' can function as a label to distinguish agents from complex physical systems, it remains controversial as to what precisely self-causation means and how the boundary should be drawn (or, how blurry the boundary is). For instance, it has been (controversially) argued that self-propelling oil droplets are agents (Hanczyc and Ikegami 2010). Consequently, many rival accounts of self-causation have been given, pointing to various factors such as internal organization, or control of environmental constraints (Moreno and Mossio 2015; Barandiaran, Di Paolo, and Rohde 2009; Skewes and Hooker 2009; Shani 2013; Burge 2009; Horibe, Hanczyc, and Ikegami 2011).[10] For the purposes of this paper we do not take a stance on how self-causation should be analyzed; what will be of importance is how it should be interpreted (i.e., whether it refers to an ontic causal process, or is a convenient heuristic).

Definition of Agential Explanation

With this operational definition of agency in place, we can introduce 'agential explanations' as scientific explanations that explain in virtue of reference to a system's agency:

10 The literature on naturalized agency is interdisciplinary to a high degree, with contributors coming from backgrounds ranging from biology or nonlinear physics to artificial intelligence, robotics, or cybernetics. A systematization of all the various contributions and approaches is still lacking.

Explanandum: In response to environmental change *E1* → *E2*, the system undergoes the change *S1* → *S2*.
Explanans: (1) The system has purpose *P*,
(2) *S2* maximizes the realization of purpose *P* in environment *E2*,
(3) the system itself is a cause of the realization of P.

An agential explanation is a type of 'extremal explanation', where the explanandum is explained as some extremal state of affairs maximizing some scalar variable *w*,[11], given certain conditions (i.e., the purpose of the system). As we will discuss later, an important class of extremal explanations, commonly used in physics, explains the explanandum as a mathematical consequence of the structural set-up of the system *S* (this typically involves various parameters *pi*).[12] By contrast, an agential explanation involves a reference to 'self-causation', where the realization of the purpose is 'caused' by the system *S* itself.

Contrast with Functional Explanation

When it comes to the use of teleological language in biology, the philosophy of science has been overwhelmingly focused on functional statements and functional explanations, e.g., "the heartbeat in vertebrates has the function of circulating blood through the organism" (Hempel 1959). Insofar as a behavior is simply an organismic trait, can one not just say that the purpose is the 'function' of purposeful behavior – thus reducing agential explanations to special cases of functional explanations?

This is not quite correct. Ascribing agency to an organism involves a different type of teleological statement than ascribing a function. The main difference between functional and agential explanations is that functional explanations attribute a purpose (function) to a *trait* of an organism, whereas agential explanations attribute a purpose to the *whole organism*. Functions are attributed to traits of organisms, whereas agency is attributed to the organisms themselves.

Nonetheless, agential and functional explanations can interact in subtle ways. For instance, a case could be made that philosophical accounts of functional explanations often presuppose it is possible to ascribe purposes to the

11 Some of the most frequently used variables include potential energy, entropy, free energy, fitness, utility.

12 See also Birch (2012) for a compatible account of what he calls 'agent-talk' in terms of robustness and stable states.

whole organism. So, for example, the causal role account of functions (roughly) holds that a function is what contributes to some 'capacity' of a larger complex system that contains it (Cummins 1975); however, how should such a 'capacity' be analyzed if not as a property of the system as a whole? Similarly, the recent organizational account (Mossio, Saborido, and Moreno 2009) uses organism-level goals that can be used to ground trait-level functions.[13]

Potentially, a similar point could be made about the selected effects account, which holds that a function is what explains why some structure was selected for in the past (Wright 1973, 1976; Millikan 1984; Neander 1991). The selected effects account presupposes there was some 'normal environment' in the evolutionary past, and while this seems like a good presupposition for structures like the heart or lungs, it is much less clear what the 'normal environment' of certain types of animal behavior should be. Since this point relates to the dispensability of agency, we will come back to this line of thought in the next section.

Agential Explanations in Social and Cognitive Sciences

In principle, agential explanations, as defined above, can also be extended to rational agents, where the purpose is defined as value or a general utility measure. Such explanations, commonplace in economics, often (and controversially) assume that economic actors are utility-maximizing agents (which is of course unrealistic, cf. Tversky and Kahneman 1974). There is a deep parallelism here between economics and behavioral ecology, noticed by Maynard-Smith in his seminal book on evolutionary game theory when he says that selection is to fitness what rationality is to economics, both being about maximization (whether of utility or fitness). This parallel also underlies formal approaches to behavioral ecology (Grafen 1984, 2014), where organisms maximize fitness in the same way rational agents maximize utility.

But the parallelism between economics and evolutionary biology goes deeper than fitness- or utility-maximization. The apparently irrational behavior diagnosed by research following Kahneman and Tverky's seminal insight on biases can be accounted for when one takes an ecological perspective. Here, considering that human agents have been shaped by evolution, and that their decision-making modules or protocols evolved in environments where information was partial and decision time was very short (due to predators, competitors etc.), then crude cognitive biases that yield a utility-enhancing solution most of the

13 For a critique, see Huneman (2019).

time would have been selected. This is what Gigerenzer calls 'ecological rationality' (Gigerenzer 2000), which gives rise to a bounded rationality, which in turn refers to how apparently irrational biases can originate as heuristics that actually are, on average, utility-maximizing given constraints (limited information and time). Thus, adopting an evolutionary viewpoint allows many instances of apparently irrational behavior to be analyzed as (boundedly) rational. All these parallels between economics and evolution by natural selection give rise to a notion of 'agency' that has recently been systematically explored (in Okasha 2018).

3 The Ontology and Dispensability of Agency

Should Agential Explanation Be Viewed Ontically?

With this systematization in place, we can now consider in some more detail the question whether the explanatory relation is to be interpreted as ontic or merely epistemic, i. e., whether an agential explanation explains in virtue of referring to an element of the ontology of the world.[14] In an agential explanation, the system itself is said to be the 'cause' of its own behavior; but what does 'causation' mean in this context?

A first safe observation is that the explanatory relation in agential explanations does not explain by simply referring to a mechanism, or to any process of causal production for that matter. There are clearly some causal processes causing the system's change of state *S1* → *S2* (e. g., neurological processes causing behavioral change), but an agential explanation, at least as stated above, does not explicitly refer to such causal processes. It explains the behavior in terms of a purpose, and a condition linking that purpose to concrete conditions in reality (i. e., *S2*). In this sense, an agential explanation cannot be viewed ontically in the same way as a causal mechanical one (Craver 2014).

A second, relatively safe observation is that if the realization of *P* is ultimately a mathematical consequence of the structural set-up of a system, then there is little reason to invoke 'purposes' and 'agents' as part of the ontology. For instance, the minimization of potential energy is a mathematically deductive consequence from the forces impinging on it as it rolls down the hill: there is no need to invoke some 'self-causation' of the rolling ball. So, if one takes 'self-cau-

14 As shorthand, one can refer to this issue as the 'ontology of agency', but just for the sake of clarity we emphasize that our approach to this issue in this paper is not directly metaphysical, but is indirect, through analyzing how agential explanations should be interpreted.

sation' to be shorthand for a pattern of behavior that is a non-causal (whether mathematical or structural) consequence of the causal set-up of the system (e. g., approach to an attractor state), then ultimately agential explanation is a non-causal explanation that identifies structural (or mathematical) consequences of how causal powers interact.

The Neo-Fisherian option largely follows this route, where organismic behavior as analyzed along the lines of the 'maximizing agent analogy', where organisms behave in such a way that maximizes their inclusive fitness. The underlying assumption is called the "phenotypic gambit" (Grafen 1984, 2014), which holds that the choice of a phenotype by the organism mirrors the allele dynamics that underlie evolution. In this way, natural selection is taken to design organisms so that they make decisions similar to what, as it were, natural selection would do if it were making the decision.

So, it would seem that an ontic interpretation of agential explanation requires a causal interpretation of agential explanation. This is indeed suggested by the inclusion of 'self-causation' in the definition of agential explanation, although the challenge for the Neo-Aristotelian option is then to specify how self-causation should be interpreted.

While Walsh is unambiguous that agents should be included in an expanded ontology (especially Walsh 2015: 211 ff.)[15], it is in his account of natural purposes that we can see how this is fleshed out. Walsh describes natural purposes as "counterfactually robust difference makers" where purpose and means are related by invariance relations (Walsh 2015: 198) in much the same way that cause and effect are related by invariance relations in Woodward's interventionist account of causation (Woodward 2003). Explaining a behavior as purposive involves identifying the disposition of "conducing" (Walsh 2015: 199), analogous to how mechanistic explanation *(sensu* Glennan 2002 or Machamer et al. 2000*)* involves identifying the dispositions of pushing or pulling (Walsh 2015: 198). Even though Walsh does not describe such relations as 'causal'[16], the account clearly involves some notion of causal difference-making where, moreover, explanations involving natural purposes are interpreted ontically.

---

15 In particular, it is implied if one adopts a Gibsonian view of the environment as a set of *affordances* proper to a species or subspecies of organism (Gibson 1979 [2014]). These affordances refer to the potential actions of an organism in an environment (e. g. running, jumping, eating, sleeping, etc.) that are jointly determined by the environment and the purposes of an agent. Moreover, these affordances in turn dispose the agent to act in certain ways.

16 In fact, at one point he emphasizes that teleological explanations are not a species of causal explanations (Walsh 2015: 196). However, here we read him as having in mind a concept of causal production.

Beyond a reluctance to expand fundamental ontology beyond what is strictly necessary, we would like to point to two reasons for being dissatisfied with the pure Neo-Aristotelian option. First, while we argued that an ontic interpretation of agential explanation should attribute some causal reality to agency, when this entails broadening the concept of causation, it becomes less clear what precisely is gained by an expanded Neo-Aristotelian ontology. Can robust patterns of counterfactual dependence be objectively judged to be causal or non-causal? This is notoriously dependent on the concept of causation one uses: once the concept is broadened enough, then any counterfactual or even counterpossible proposition will appear as causal (Huneman 2010). After all, if the agent with purpose *P* were modified to an agent with purpose *P'*, then the observed behavior would be (much) less likely, and this is sufficient to count as a causal relation for some accounts of causation. This raises the question: if agential explanation ultimately boils down to patterns of counterfactual relations, why does it matter if one interprets agential explanation ontically?

Second, pragmatic factors complicate the ontic interpretation of agential explanation. An agential explanation entails some counterfactual relation between explanans (agent, purpose *P*) and explanandum (behavior); however, both a particular explanandum behavior as well as the agent's purpose *P* can be described at finer and coarser grains. Depending on the granularity with which explanandum/explanans is described, the causal character of the corresponding explanatory relation changes (for an argument, see Desmond 2019). If agential explanation is to be viewed ontically, and thus as explanatory in virtue of picking out some self-causing capacity of an organism, one would not want this causal character to disappear merely due to pragmatic factors, such as the grain at which explanandum/explanans is described.

Can Agency Be Dispensed with?

While we do not pretend to have given any direct argument against the ontic interpretation of agential explanation, we do hope to have clarified how the ontic interpretation will lead to a host of problems, some of which are perhaps insuperable. However, we now want to turn our attention to the other side of the coin: explanatory dispensability. The strongest argument in favor of a non-ontic conception of agential explanation is that agency is dispensable (this is

Ockham's razor).[17] We will illustrate in this section how it is misguided to believe that organismic agency has been dispensed with in science. Whether science may be able to dispense with agency in the future is a whole other question. What we wish to show is a more limited point: given a pessimistic meta-induction on attempts to dispense with agency, there are at least good grounds to believe that science will *not* be able to dispense with agency. This will lay the ground for the Kantian approach, which uniquely combines two positions: agency is indispensable (for scientific reasons), but agential explanations should not be conceived ontically.

Many phenomena clearly do not ask for agential explanations. If a tree branch cracks and falls to the ground during a storm, and we seek to explain the change in the tree's state, we spontaneously tend not to appeal to any type of 'agency' of the tree. A property of the tree as a whole could be explanatorily relevant – for example, a disposition such as brittleness could be referred to in order to explain why oaks tend to crack more than willows during storms. Nonetheless, we tend not to explain this tree 'behavior' in terms of the purposes of the tree. Rather, given certain forces created by the wind, and perhaps given certain structural properties of the tree, the outcome of the branch cracking was determined. No agency is involved.

Even if extremal explanations were to be used, no agency would be required. Classical mechanics provides a perfect example of how extremal explanations exist alongside causal-mechanical explanations. Newtonian analyses of the behavior of masses, in terms of a mechanistic account of the continued action of local forces, can always be rephrased with the Principle of Least Action (through the Hamiltonian or Lagrangian formalism), which abstracts away from a great number of degrees of freedom in a system, and instead ascribes a certain scalar (i.e., the 'action' S) to a system. The behavior of the system is then the behavior that maximizes or minimizes the action (cf. Coopersmith 2017).[18]

When a system has an extremely large number of degrees of freedom ($\sim 10^{23}$), a different type of extremal explanation is needed, but even here the explanation remains non-agential. Consider a generic thermodynamic phenomenon, such as

17 Going in the other direction, from indispensability to a realist interpretation of a concept, is more controversial but has of course been widely explored since Quine and Putnam.

18 Historically, such non-agential extremal explanations are exactly what Leibniz had in mind, when he stated that each mechanistic explanation, given in terms of the differential equations governing the trajectories of the parts, could be reformulated in terms of final causes (*Discours de métaphysique* § 13 (Leibniz, 1890)). As a further aside, even though such final causes were dispensable for Leibniz, explanations involving them were to be preferred because they were the most conducive to theology and were compatible with God's moral maxims.

the flow of heat from hot to cold. In statistical-mechanical analysis, the molecules in a gas or liquid fluctuate randomly, but after some time, it is likely that the faster-moving molecules will not remain bunched up in one area of the container (i. e., the 'hot' area), but will spread out over the whole container, either by diffusion or by transferring momentum to slower-moving molecules through collisions.

This type of explanation, first introduced by Boltzmann, is non-causal in the sense that it relies only on principles of combinatorics together with some boundary conditions. A uniform temperature is a vastly more likely outcome than any other since it corresponds to a much greater number of possible microstates, or ways of distributing molecular speeds among the molecules in the container. Erwin Schrödinger aptly named this type of extremal explanation, the 'method of the most probable distribution' (Schrödinger [1946] 2013).

This type of extremal explanation has been widely applied to more complex systems, including open systems that are far from thermodynamic equilibrium ('dissipative systems'). Ilya Prigogine, a pioneer in this field, proposed the principle of minimal entropy production: i. e., systems in far-from-equilibrium conditions organize themselves so as to minimize the increase of entropy (Prigogine 1947). However, universal extremal principles that govern the behavior of all dissipative systems have not been found. For instance, the principle of maximal entropy production has also been proposed (Paltridge 1979). It remains unclear to what extent these extremal principles are instances of the method of the most probable distribution – or whether they bring goals and purposes to the table that cannot be explained through statistics alone.

When we look at more recent applications of statistical physics, gradual progress can be discerned. For instance, an upper bound on the rate of bacterial replication has been proposed (England 2013). However, this remains a research program, and while there is not yet any clear reason why the program cannot continue to make gradual progress, the prospect of reducing organismic behavior to statistical physics remains remote.

When one departs from the reductive rigor of statistical physics, then axiomatic thermodynamic extremal principles can seemingly be used to explain animal behavior (and human behavior in particular). The work of Karl Friston, which has enjoyed success in theoretical neuroscience, is an example of this approach. Here animal behavior is analyzed as minimizing free energy – intuitively, this means that organisms minimize the quantity of 'surprise' in their environment (Friston 2010). However, free energy minimization is taken as axiomatic and is not given a deeper derivation in the way Boltzmann had done for entropy maximization in the context of equilibrium thermodynamics. In this way, it

seems that the concept of 'organismic purpose' (in this case, the purpose of minimizing free energy) cannot be dispensed with within Friston's framework.

We have discussed two types of non-agential extremal explanations – causal mechanical ones, and non-causal statistical ones – and argued that both fail to dispense with organismic agency. We would now like to consider in more detail what is perhaps the most serious contender for dispensing with organismic agency, namely selectionist explanations. To what extent selectionist explanations reduce to non-causal statistical ones remains controversial. Some have argued that they do: evolution caused by fitness differences is structurally identical to, for instance, the differential growth rates of bank accounts with different interest rates (Matthen and Ariew 2002; Walsh, Lewens, and Ariew 2002).

Regardless of the interpretation of natural selection, it is clear how it should be combined with causal-mechanical explanation so as to seem to dispense with agency. Consider the behavior of chemotaxis, where bacteria swim up sucrose gradients. An agential explanation of this behavior would refer to the purpose of the bacteria to get nutrition. However, one could attempt to explain chemotaxis by referring to how, given a certain environmental input into the mechanism of chemotaxis, a certain output (swimming behavior) is to be expected. And why is the mechanism of chemotaxis set up in this particular way (connecting these inputs with those outputs)? Here the selectionist explanation comes in: those bacteria that came up with the mechanism of chemotaxis were able to take distance from competitors and maximize their access to resources (Wei et al. 2011). This in turn allowed them to reproduce more, eventually crowding out the bacteria incapable of chemotaxis. There is no need to reference agency here.

Can agential explanation be reduced to selectionist explanation in this way? We will not take a stand on whether it can or not ; however, we would like to argue that this issue is considerably more complicated than the simple selectionist-mechanical explanation above suggests. A selectionist explanation may be adequate for chemotaxis, but it is far from clear that this can be generalized to organismic behavior in general, especially concerning cases where organisms produce adaptive behavior even in novel environments.

To see this, recall that a selectionist explanation requires a homogeneous selective environment (Brandon 1990), which means that selection pressures must be relatively uniform across the environment. So, if an organism is exposed to a 'novel' selective environment, this means that it is exposed to selection pressures that the organism's ancestors never encountered. One of two scenarios then presents itself. The first is that the fitness-maximizing analogy breaks down: the organism sticks to its behavior that was previously adaptive, but maladaptive in the new environment. The second is that the fitness-maximizing analogy holds, and the organism chooses a new behavior that maximizes its (in-

clusive) fitness. However, in this case, referring merely to a selectively-determined function cannot explain why the new behavior was chosen. In other words, a mere selectionist explanation is not adequate.

To give this line of thought some more systematic detail: assume that organism *O*'s behavior *B* was selected for in selective environment *E*. Furthermore, *B* is the output of some heritable function *F*. What if the environment shifts to some radically different *E*?* Different organisms will behave differently, depending on *F*. Some will continue producing *B* regardless; others will be sensitive to cues in the environment, and the function *F* will produce as output *B** instead of *B* (this is behavioral plasticity). If *B** turns out to be adaptive to *E**, is this not a lucky coincidence assuming that *O*'s ancestors never encountered the selection pressures in *E*?* In other words, if *F* is designed for *E,* is it not a lucky coincidence that *F* should also produce adaptive behavior in a radically different environment? This is how organismic purposes and agency can be introduced, to provide a better explanation of the production of adaptive behavior in novel environments.

Of course, plasticity itself can be selected for (Bradshaw 1965), and this is where the problem gets complex and interesting. So, if *F* underlies a plastic trait that is modulated to produce adaptive behavior in *E**, the selectionist could respond that the appropriate explanation is not an agential explanation, but rather that *E** and *E* are simply not two different selective environments. They may be different *physical* environments, but they are instances of the same selective environment – for instance, they may be similar instantiations of the same pattern of heterogeneity, such as possessing the same varying cues.

In this way, the question of whether agential explanations can be reduced to selectionist explanations opens up to a large and fundamental problem of what selective environments are and how they should be delineated. For this reason, we do not wish to take a stand on whether agential explanations can be reduced to selective explanations. A safer conclusion we would like to draw is this: it is currently unclear whether agential explanations can be reduced to selectionist explanations, and therefore we should not assume that the theory of natural selection easily dispenses with agency. We should take seriously the option that agency may be indispensable.

4 The Kantian Approach to Purposiveness

Kant's work on teleology can offer an interesting perspective in that he considered a closely related problem – apparently incompatible ways of viewing biological organisms – but resolved it in a way that cuts across the dichotomy

that pairs the indispensability of agency with an ontic view of agential explanation. Most interest in the Kantian perspective on teleology has focused on developmental phenomena.[19] A passage that is often quoted as particularly relevant is the following where Kant introduces the term 'self-organization':

> In such a product of nature each part is conceived as if it exists only *through* all the others, thus as if existing *for the sake of the others* and *on account of* the whole, i.e., as an instrument (organ), which is, however, not sufficient (for it could also be an instrument of art, and thus represented as possible at all only as a purpose); rather it must be thought of as an organ that *produces* the other parts (consequently each produces the others reciprocally), which cannot be the case in any instrument of art, but only of nature, which provides all the matter for instruments (even those of art): only then and on that account can such a product, as an *organized* and *self-organizing* being, be called a *natural purpose*. (Original emphasis; translation slightly modified. Kant [1790] 2001, 274; 5:374)

What Kant is arguing here is that organisms are not simply machines (e.g., artifacts), where each part may be designed to contribute to the whole, but where some external source (e.g., the artisan) is the cause of the production and maintenance of each part of the machine. Thus, for instance, the minute hand of a watch is produced by the artisan and not by any other part of the watch. By contrast, the various anatomical and physiological traits of an organism are produced by processes internal to the organism. Organisms are thus not to be judged as machines: an essential property of organisms is that the parts also cause the production and maintenance of the other parts, as we see in the ontogenesis.

In this way, the passage in which Kant introduces the notion of self-organization is most directly relevant to issues concerning the development of organisms; not surprisingly this is where the connection between contemporary biology and Kant's thought has most often been made (see also Huneman 2017). Here however, we would like to draw out more explicitly the implications of this for organismic behavior and organismic agency.[20] In particular, we will look in more detail at Kant's general idea of purposiveness, and at his general

19 For a discussion of these different perspectives, and the relevance of the Kantian approach to contemporary debates in evolutionary biology, see Huneman 2017. See also references to Kant in Varela 1979, Kauffman 1993.

20 To a certain extent, the division between development and behavior is artificial. Development typically refers to morphological changes (cell differentiation, growth, etc.) that are relatively irreversible and slow in comparison to physiological changes (metabolism) or behavioral changes (movement through space). Some explicitly distinguish between development and behavior (e.g. Burge 2009); by contrast, most behavioral ecologists consider any trait (for instance a tree growing small vs. large leaves) as a 'behavior'.

treatment of the antinomy of teleological judgment, which concerns the apparent clash between 'mechanistic' and 'teleological' approaches to the organism.

4.1 The Antinomy of Teleological Judgment

In his *Critique of the Power of Judgment,* Kant posits the following two conflicting maxims concerning 'generation' (a contemporary close-equivalent: development) and 'mechanical laws' (namely, laws that govern the way parts yield wholes – see McLaughlin 1990):

Thesis: All generation of material things is possible in accordance with merely mechanical laws.

Antithesis: Some generation of such things is not possible in accordance with merely mechanical laws. (Kant [1790] 2001, 258 – 259; 5:387).

In particular, Kant had biological organisms in mind as possible entities that are not generated merely according to mechanical laws. This thesis-antithesis pair is simply a contradiction, leading to mutually incompatible views with no prospect of reconciliation.

Kant's first step, then, is to make explicit that such pronouncements about the nature of reality are actually *judgments* that are necessarily relative to our *cognition* of reality. Hence, he proposes the following thesis-antithesis pair:

> The *first maxim* of the power of judgement is the *thesis:* All generation of material things and their forms must be judged as possible in accordance with merely mechanical laws. The *second maxim* is the *antithesis:* Some products of material nature cannot be judged as possible according to merely mechanical laws (judging them requires an entirely different law of causality, namely that of final causes). (Kant [1790] 2001, 258 – 259; 5:387)

This is the antinomy of teleological judgment. The motivation underlying the antithesis draws on the idea that mechanical laws do not seem to adequately account for the organization that can be found in biological organisms. In particular, Kant writes:

Nature, considered as a mere mechanism, could have formed itself in a thousand different ways without hitting precisely upon the unity (KU, AA, V: 360).

The mechanical laws do not privilege any particular organization over another; hence, if the organization were to be explained with merely mechanical laws, the organization of organisms could only be judged to be the result of *chance* (see Huneman 2006).

4.2 Contingency and Kant's Concept of Purposiveness

In this way, Kant is relying on a concept of purposiveness that can be described as the 'lawfulness of the contingent as such' (First Introduction to the *Critique of Judgement,* Kant [1790] 2001, 20; 20:217). An initial illustration of the concept can be given in the context of development. For instance, if one were only to take the mechanical laws of nature into account, the fact that the development of a chicken leads to a chicken appears to be contingent – once the initial and boundary conditions are sufficiently changed, it might develop into a monster. However, the laws themselves cannot explain this wide divergence in outcome, since in both cases the same laws apply. One must introduce the idea that the development is the development *of a chicken,* and therefore is oriented towards this goal. Thus, such an idea brings some necessity into a process that is, with regard to nature itself (i.e., the mechanical laws of nature), contingent. The same goes for the functions of organisms: whether or not an elephant's lungs breathe seems highly contingent if one only takes into account the laws of nature, but appears as necessary when we introduce the idea that breathing is the function of the lung. This entails invoking the idea of a functioning organism. Thus, biological functions and embryogenic development instantiate the same epistemic pattern (Huneman 2006).

Kant's theory of purposiveness is intended to reflect this epistemic fact. To introduce it, he gives a famous example: what if one were to come across a regular hexagon drawn in the sand (Kant [1790] 2001 §62)?[21] This, says Kant, can only be understood as an instance of purposiveness, because if we do not posit a concept ('regular hexagon') that is 'at the basis of' (i.e., guided) its production, we cannot understand why it is drawn in the sand. In other words, while the laws of nature can lead to the appearance of all sorts of figures in the sand, the specific kind of figure we see is not privileged by those laws (i.e., it is not any more probable than any other kind of figure). When we see a physical instance of a regular hexagon then, and we judge that it fits the concept of a 'regular hexagon', there is no indication in the laws of nature as to why a regular hexagon should be produced rather than another one. Hence, we reasonably assume that the concept 'regular hexagon' was at the basis of its production – namely, someone thought of this concept and has drawn the hexagon – and thereby

21 Note that Kant chose an example from mathematics as part of his overall strategy to decouple the notion of purposiveness he intends to capture from the usual scheme of craftsmanship, fabrication, etc.

the contingent figure we see on the sand features some lawlikeness (since it has been drawn according to some rule).

To put the argument in a more contemporary idiom, consider the following. Among the set of all possible hexagons, the size of the subset of regular hexagons (i.e., with equal sides) is extremely small (measure = 0). Hence, given that we observe a regular hexagon, and that the probability that a process governed only by mechanical laws of motion would cause a regular hexagon to appear is 0, appealing to the presence of a concept at the ground of the production allows for a (much) better explanation of the appearance of the hexagon.

This same line of reasoning can be applied to organisms. In Kant's *Unique Argument for a proof of God's existence*, the first major text in which he deals with life and finality, he considers the traditional example of the eye, describing the example in the following way: in an eye there are many parts, each following different and mutually independent laws, and yet, the parts function not only in such a way that the eye can see, but if we were to even slightly change the structure or behavior of one of the parts, the eye as a whole would no longer achieve sight. Similarly, in the third *Critique* Kant gives the example of the bird whose different anatomical parts seem to be organized in very specific ways in order to enable flight: "the structure of a bird, the hollowness of its bones, the placement of its wings for movement and of its tail for steering, etc." (Kant [1790] 2001, 233; 5:360). And yet, it remains possible also to view an organism as a clump of dead matter, obeying the laws of mechanics. The price to pay for the latter possibility is that there is no longer any answer to the question of why those parts are so *contrived* – to use a word that will become crucial for Darwin – to allow flight.[22]

The notion of purposiveness as elaborated by Kant is closely related to extremal explanations. Consider the evolution of the camera eye, and its dependence on the laws of nature (and causal-mechanical processes). Assume we can vary the laws of nature (and causal-mechanical processes) by manipulating a parameter vector $\{(Ai)\}$, and let some scalar variable W represent the functional

22 Note that the judgment that there is a causal relation between two objects or events (like two billiard balls colliding) is a constitutive use of reason (understanding), whereas judging according to mechanical laws is a regulative use of reason, even though mechanical laws are clearly closely related to causality as an *a priori* principle. But, as said before, mechanism is about the relation between parts and wholes – knowing wholes from the parts – while causation is about the succession of events or facts. Disentangling how precisely Kant understood the relation between causality and mechanical laws is the subject of some debate in Kant scholarship. See, for example, Allison 2001.

value of the eye (for instance, the representational accuracy[23] of its images). Further assume that a particular set of parameter values (*A1, A2... An*) maps onto the extreme value *W0* for *W* – namely, the best functionality or representational validity – and let *W0* also be the precise value that *W* assumes in empirical nature. Yet, it would seem highly improbable that the parameter vector should attain the exact (*A1, A2... An*) among all possible values of {(*Ai*)}; correspondingly, functional sight seems highly improbable. Referring to the concept of 'sight' as a concept that somehow guides the fixation of the parameter values (*A1, A2... An*) allows one to ascribe a kind of necessity (or at least a much higher probability) to (*A1, A2... An*). Thus, interpreting *W0* as the purpose of 'sight', allows for the explanation of why sight-enabling structures emerged.

So, when Kant holds here that the fact that (A_1, A_2... A_n) obtains is not explainable except if one thinks of a concept ('sight') at its ground, this account is perfectly analyzable as positing an extremal explanation which explains the explanandum as the extremal value (in turn corresponding to functional sight) of a mathematical function. Thus, the reason why the vector (A_1, A_2... A_n) – otherwise wholly contingent – is the one that we find in nature is that the vector realizes some extremum. Moreover, the concept of vision ultimately appeals to the idea of a functioning organism that is able to survive (e.g. catch prey, avoid predators, track motions and light in its environment, etc.). In this way, the reference to the concept of vision introduces lawlikeness into the contingent unity of mechanical laws involved in the design of the eye.

The lawfulness of this contingent unity, i.e., the notion of purposiveness, is for Kant only a *regulative* and not a *constitutive* concept or principle. Regulative and constitutive principles refer to two uses of reason. While not going into too much detail,[24] the latter refer to the synthetic *a priori* principles (causality, permanence, reciprocal action, etc.) that ground any science of nature, and when events or facts can be subsumed under such principles, they can be considered as 'objective'. By contrast, in the regulative use of reason, the principles inform our cognition of the objects and allow for knowledge of objects, but do not posit anything as objective. An instance of the "regulative use" of ideas of reason, described in the *Dialectic* or the *Critique of Pure Reason*, involves prescribing the idea of the "synthesis of all conditions" to the world. This allows us to require new conditions for the conditioned events, empirical laws, forces or facts we have found. Nonetheless, we cannot posit as objective the *whole* of conditions –

23 See Burge (2009) on veridicality as a norm for representational systems. Burge's notion of norm is here accounted for in terms of extremal value.

24 For an introduction see Huneman 2007.

which Kant calls the "unconditioned" and which can refer to, for instance, the whole world, or God (which in turn refers to what Kant calls "Ideas of reason"[25]). Likewise, the idea that each individual belongs to a species that in turn belongs to a higher order class (family, genus, etc.) is not an objective fact, but a regulative principle of our knowledge, without which we would be unable to cognize an ordered world.

The regulative principles that allow for biology (since, at least as stated in the third *Critique*, the constitutive principles of judgment lead to viewing the contingent as simply contingent and lawful) are precisely the lawlikeness of the contingent as purposive: this kind of lawlikeness implies, as we said, the idea of a functional or developing organism. Moreover, from the moment the reference to such totality – namely, the organism – is introduced, a new level of necessity is brought into a set of facts and events that would otherwise appear wholly contingent. This then allows these facts to be studied in a scientific manner: biologists will ask which mechanisms fulfill this or that function, or what processes lead to the formation of such and such an organ and then the whole organism. Inversely, any such scientific enquiry already assumes the lawlikeness of the contingent.

5 Organismic Agency and the Demand of Reason

We will now depart from a description of Kant's framework, and draw out the (Kant*ian*) implications of the concept of purposiveness for the central issue of this paper, namely the ontology and dispensability of agency. In particular, we will argue against three ways of viewing agency: first the position that agency is a mere projection (non-ontic, dispensable); second that it refers to an element of objective nature (ontic, indispensable); finally, we will also contrast the Kantian view with the position that agency is a mere heuristic, but that it is indispensable given our evolved nature and limited computational capacity. This is a view where agential explanation is viewed as non-ontic, and agency as dispensable for cognition of reality, but indispensable for the *human* cognition of reality.

First, can purposiveness be seen as a projection of the human mind onto the natural world? In this view, goals and functions are in fact anthropomorphic projections onto the world (e. g., Lewens 2007: 544 – 5). Such projections may serve some purpose as heuristics, but they do not reveal anything objectively real

25 On this notion see Allison 2001, Grier 1995.

about the world and are entirely dispensable: they are to be replaced by mechanistic or law-based explanations whenever the latter become available.

In response, recall from the previous section how, in a Kantian framework, the question whether or not agential explanation should be viewed ontically is bound up with the distinction between 'regulative' and 'constitutive' principles. Only the latter give rise to ontic explanations; nonetheless, that does not mean that purposiveness is merely a projection of the human mind onto the natural world. Granted, purposiveness does not constitute nature as such and is therefore not objective in the same way that laws of nature are. However, they are not a 'projection' in the sense that it is an optional way for a cognizing subject to see the world. Once biological items are the object of a quest for knowledge, there is no alternative to purposiveness for the faculty of knowledge.

This can be emphasized by referring to one last element in Kant's work, namely how the structure of the faculty of knowledge is 'finite':

> Absolutely no human reason (or even any *finite reason that is similar to ours in quality, no matter how much it exceeds it in degree*) can ever hope to understand the generation of even a little blade of grass from merely mechanical causes. (our emphasis, Kant [1790] 2001, 279; 5:410)

Our reason is 'finite' because it cannot derive intuitions from concepts, and therefore, the particular from the universal.[26] An 'infinite' reason, by contrast, would not be limited in this way. However, Kant does not posit that such an infinite reason actually exists, or for that matter, is even possible under some counterfactual scenario. Instead, it is a mere idea that orients philosophical enquiry into knowledge, or if you will, a thought experiment aimed at clarifying what reason is. This distinction between finite and infinite reason can be connected to two modes of understanding: discursive and intuitive understanding. Intuitive understanding (which, like infinite reason, is a mere idea that orients the philosophical enquiry about knowledge) would be able to cognize the particular instances of concept X at the same time it cognizes the (universal) concept of X. By contrast, discursive understanding must go through 'mediations' in order to arrive at the particular. Simple acts of observation can be such mediations (to check whether anything corresponds to the concept X).

The concept of purposiveness is also such a mediation, since it allows reason to proceed from the universal laws of nature to particular organisms. A living being can, in general, be analyzed by means of mechanistic laws, e. g., the uni-

26 Concepts allow us access to the universal, while intuitions provide us access to the particular.

versal laws governing the dynamics of each part. However, here only a very specific combination of the part-level processes results in a living being (think of the various laws involved in the building of the eye, mentioned above). Hence, the finite reason has to shift to the level of the "lawlikeness of the contingent as such", namely, to assume the regulative principle of purposiveness by introducing the reference to the whole organism. This "idea of the whole," he says in § 65, is only a principle of cognition, not of production.

This finiteness of reason leads to what Kant in other passages describes as a 'demand' of reason for the 'unconditioned'. We previously described it as the 'synthesis of all conditions', but in a more contemporary idiom, it could also be described as the following:

> The demand for the unconditioned is essentially a demand for ultimate explanation, and links up with the rational prescription to secure systematic unity and completeness of knowledge. Reason, in short, is in the business of ultimately accounting for all things. (...) the demand for the unconditioned is inherent in the very nature of our reason, [and] is unavoidable and indispensably necessary... (Grier 2018)

Kant thus takes this demand of reason to deliver a kind of impossibility result for the possibility of a non-purposive explanation of organismic development (and by extension, the same could be said of organismic agential-like behavior).

In this way, in contrast to interpreting agential explanations as involving anthropomorphic projection, for the 'Kantian option' there is no alternative to explaining organismic behavior as agential. Moreover, seeing an organism as an agent is even a precondition (the transcendental ground) to being able to make a projection onto a natural system. For instance, if in some agential explanation, a repertoire of actions is projected onto a living organism, this presupposes seeing an organism as an agent. Assuming agency makes ascribing empirical methodology and even (behavioral) property to organisms possible. This is how the 'indispensability' implied by the Kantian option should be understood.

Others have taken the 'blade of grass' passage cited above as support for an ontic view of agential explanation, where "organisms are subjects having purposes according to values encountered in the making of their living" (Weber and Varela 2002, 102). But an ontic interpretation of agential explanation – where organisms are (objectively) subjects with (objective) purposes – clashes with Kant's overarching transcendental framework, since only constitutive principles can ground ontic explanations. Given that regulative principles such as purposiveness are a consequence of the finite nature of reason, they are not empirically discoverable facts, but are instead presupposed in any epistemic strategy for searching empirical truths. This shows how the Kantian option implies a non-ontic view of agential explanation.

Does the Kantian option imply an epistemic view of agential explanation? We take 'epistemic' here to refer to expectability *sensu* Salmon (Salmon 1989), where an explanation explains in virtue of showing the explanandum as expected (i. e., with high probability). In this sense, the Kantian option does certainly interpret agential explanations as showing how the explanandum is to be expected; however, much also depends on how 'expectability' is interpreted. Consider the subjective interpretation[27], where expectability is analyzed as dependent on the amount of information available to the subject; as the information changes, so does the expectability. This is what is presupposed if one views agential explanations as arising from bounded rationality, where agency is ascribed as a heuristic or computational shortcut given time and/or information constraints. The Kantian option is not 'epistemic' in this way: it does not refer to properties of what could be called 'evolved human nature' but rather to a fundamental structure of reason itself. Any finite reason, even if it would be as computationally powerful as the largest supercomputer, would not be able to understand organismic purposes only in terms of causal mechanisms. Even if our empirical nature were very different – for instance, if we had evolved very different cognitive heuristics for understanding the world – as long as we are endowed with a finite reason then we would still employ teleological concepts such as agency. Thus, agential explanations are non-ontic in the sense that agency as a concept ultimately can be traced back to a fundamental structure of reason (and not a structure of the objective world, nor to a quantity of information about the world available to a subject).

In sum, the Kantian approach suggests that agential explanations are to be viewed as non-ontic explanations but in which agency is indispensable. Viewing organisms as agents is a heuristic – it allows organisms to be identified as wholes in the first place (cf. Breitenbach 2008), and thereby allows a research program about the mechanisms of functions and development – but it is not merely a heuristic: it is unavoidable for a *finitely rational understanding* of nature. Agential explanations may be predictive tools – they may accurately summarize complex patterns of behavior and allow us to predict how organisms will respond to environmental inputs – but they are more than mere predictive tools, because if they were merely predictive tools, agential explanations would be replaceable by an explanation that integrates a mass of complex causal detail. Even though the latter may be predictively equivalent or even superior to an agential explanation, it does not afford *understanding* to rational beings.

---

27 An objective interpretation seeks to analyze expectability (and probability) in terms of objective structures, and thus leads to a variation on the ontic view of agential explanation.

6 Conclusion: Organismic Agency and the Demand of Reason

The shift in contemporary biology towards the agential approach motivates paying closer philosophical attention to agential explanations. Yet agential explanations are still today interpreted along the lines of a dichotomy between ontic/indispensable or non-ontic/dispensable, even though both options are ultimately unsatisfactory. In this paper we elucidated the Kantian option, where viewing organisms as agents is a demand of reason, and thus indispensable to our cognition of reality, but yet where agency is not added to the 'furniture' or basic ontology of the world.

This implies that agential explanations are *unavoidable* given our rational nature. This goes further than merely stating that agential and non-agential explanations are complementary. While it is of course possible also to view organisms as combinations of mechanisms, scientists, as rational beings, have no choice but to use agential explanations as well. Agency is thus not simply an investigative heuristic or a predictive tool that can be dispensed with once our scientific knowledge is sufficiently advanced, like a ladder that is climbed only then to be kicked away. Seeing agency in the natural world is not like a form of superstition that can be dispelled by the onward march of scientific reason; it is inherent to reason itself and is therefore not a ladder that can ever be kicked away.

References

Allison, Henry E. (2001): *Kant's Theory of Taste: A Reading of the Critique of Aesthetic Judgment.* Cambridge, UK: Cambridge University Press.

Atran, Scott (2002): *In Gods We Trust: The Evolutionary Landscape of Religion.* Oxford, UK: Oxford University Press.

Auletta, Gennaro (2013): "Information and Metabolism in Bacterial Chemotaxis." In: *Entropy* 15, pp. 311–326.

Barandiaran, Xabier E./ Di Paolo, Ezequiel/ Rohde, Marieke (2009): "Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action." In: *Adaptive Behavior* 17, pp. 367–386.

Barrett, Justin L. (2000): "Exploring the Natural Foundations of Religion." In: *Trends in Cognitive Sciences* 4, pp. 29–34.

Bateson, Patrick (2005): "The Return of the Whole Organism." *Journal of Biosciences* 30 (1), pp. 31–39. https://doi.org/10.1007/BF0270514.

Birch, Jonathan (2012): "Robust Processes and Teleological Language." In: *European Journal for Philosophy of Science* 2, pp. 299–312.

Bradshaw, Anthony D. (1965): "Evolutionary Significance of Phenotypic Plasticity in Plants." In: *Advances in Genetics* 13, pp. 115 – 155.

Breitenbach, Angela (2008): "Two Views on Nature: A Solution to Kant's Antinomy of Mechanism and Teleology." In: *British Journal for the History of Philosophy* 16, pp. 351 – 369.

Burge, Tyler (2009): "Primitive Agency and Natural Norms*." In: *Philosophy and Phenomenological Research* 79, pp. 251 – 278.

Brandon, Robert N. (1990):wei *Adaptation and Environment*. Princeton University Press.

Calvo Garzón, Paco/Keijzer, Fred (2011): "Plants: Adaptive Behavior, Root-Brains, and Minimal Cognition." *Adaptive Behavior* 19, pp. 155 – 171.

Caporael, Linnda R./Griesemer, James R./Wimsatt, William C. (2013): *Developing Scaffolds in Evolution, Culture, and Cognition*. Cambridge, MA: MIT Press.

Coopersmith, Jennifer. (2017): *The lazy universe: an introduction to the principle of least action*. Oxford, UK: Oxford University Press.

Craver, Carl F. (2014): "The Ontic Account of Scientific Explanation." In: Kaiser, Marie I./ Scholz, Oliver R./Plenge, Daniel/Hüttemann, Andreas (Eds.): *Explanation in the Special Sciences*, Dordrecht: Springer Netherlands, pp. 27 – 52.

Cummins, Robert (1975): "Functional Analysis." *The Journal of Philosophy* 72 (20): 741 – 765.

Dawkins, Richard (1976): *The Selfish Gene*. Oxford, UK: Oxford University Press.

Dawkins, Richard (1982): *The Extended Phenotype: The Long Reach of the Gene*. Oxford, UK: Oxford University Press.

Desmond, Hugh. (2018). Natural selection, plasticity, and the rationale for largest-scale trends. In: *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *68 – 69*, 25 – 33.

Desmond, Hugh. (2019). Shades of Grey: Granularity, Pragmatics, and Non-Causal Explanation. In: *Perspectives on Science*.

Duijn, Marc van/Keijzer, Fred/Franken, Daan (2006): "Principles of Minimal Cognition: Casting Cognition as Sensorimotor Coordination." In: *Adaptive Behavior* 14, pp. 157 – 170.

England, Jeremy L. (2013): "Statistical Physics of Self-Replication." In: *The Journal of Chemical Physics* 139, 121923.

Fisher, Ronald A. (1919): "The Correlation between Relatives on the Supposition of Mendelian Inheritance." In: *Transactions of the Royal Society of Edinburgh* 52, pp. 399 – 433.

Fisher, Ronald A. (1930). *The Genetical Theory of Natural Selection*. Oxford, UK: Oxford University Press.

Friston, Karl. (2010): "The Free-Energy Principle: A Unified Brain Theory?" In: *Nature Reviews Neuroscience* 11, pp. 127 – 138.

Gibson, James J. ([1979] 2014). *The Ecological Approach to Visual Perception*. New York and London: Psychology Press.

Gigerenzer, Gerd (2000): *Adaptive Thinking: Rationality in the Real World*. Oxford, UK: Oxford University Press.

Glennan, Stuart (2002): "Rethinking Mechanistic Explanation." In: *Philosophy of Science*, 69 pp. 342 – 353.

Godfrey-Smith, Peter (1996): *Complexity and the Function of Mind in Nature*. Cambridge, UK: Cambridge University Press.

Grafen, Alan (1984): "Natural Selection, Kin Selection and Group Selection." In: Krebs, John Richard/Davies, Nicholas Barry (Eds.): *Behavioural Ecology*. Oxford, UK: Blackwell, pp. 62–84.

Grafen, Alan (2006): Optimization of inclusive fitness. In: *Journal of Theoretical Biology* 238, pp. 541–563.

Grafen, Alan (2014): "The Formal Darwinism Project in Outline." *Biology & Philosophy* 29, pp. 155–174.

Grier, Michelle (1995): "Kant's Rejection of Rational Theology." In: *Proceedings of the Eighth International Kant Congress* 2, pp. 641–650.

Grier, Michelle (2018): "Kant's Critique of Metaphysics." In: Zalta, Edward N. (Ed.): *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2018/entries/kant-metaphysics/.

Hamilton, Matthew B. (2009): *Population Genetics*. Hoboken, NJ: Wiley-Blackwell.

Hanczyc, Martin M./Ikegami, Takashi (2010): "Chemical Basis for Minimal Cognition." In: *Artificial Life* 16, pp. 233–243.

Hempel, Carl G. (1959): "The Logic of Functional Analysis." In: Gross, Llewellyn (Ed.): *Symposium on Sociological Theory*. New York: Harper and Row, pp. 271–87.

Horibe, Naoto/Hanczyc, Martin M./Ikegami, Takashi (2011): "Mode Switching and Collective Behavior in Chemical Oil Droplets." In: *Entropy* 13, pp. 709–19.

Huneman, Philippe (2006): "Naturalising Purpose: From Comparative Anatomy to the 'Adventure of Reason.'" In: *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 37, pp. 649–674.

Huneman, Philippe (2007): "Reflexive judgement and Wolffian embryology: Kant's shift between the first and the third *Critique*." In: Huneman, Philippe (Ed.): *Understanding purpose? Kant and the philosophy of biology*. Rochester: University of Rochester Press, pp. 75–100.

Huneman, Philippe (2010): "Topological Explanations and Robustness in Biological Sciences." In: *Synthese* 177 (2), pp. 213–45. https://doi.org/10.1007/s11229–010–9842-z.

Huneman, Philippe (2017): "Kant's Concept of Organism Revisited: A Framework for a Possible Synthesis between Developmentalism and Adaptationism?" In: *The Monist*, 100, pp. 373–390.

Huneman, Philippe (2019): "Revisiting Darwinian Teleology." In: *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 76: 101188.

Huxley, Julian ([1942] 1974): *Evolution: The Modern Synthesis*. London, UK: Allen and Unwin.

Jacob, François/Monod, Jacques. (1961). "Genetic regulatory mechanisms in the synthesis of proteins." In: *Journal of Molecular Biology* 3, pp. 318–356.

Kant, Immanuel ([1790] 2001): *Critique of the Power of Judgment*. Cambridge, UK: Cambridge University Press.

Kauffman, Stuart A. (1993): *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford, UK: Oxford University Press.

Leibniz, Gottfried W. (1890): "Discours de métaphysique". In: *Philosophische Schriften* Band IV, Carl Immanuel Gerhardt, (Ed.), pp. 427–465.

Lewens, Tim (2005): *Organisms and Artifacts: Design in Nature and Elsewhere*. Cambridge, MA: MIT Press.

Lewens, Tim (2007): "Function." In: Matthen, Mohan/Stephens, Christopher (Eds.): *Handbook of the Philosophy of Science*. Amsterdam: Elsevier, pp. 525–47.
Lyon, Pamela (2015): "The Cognitive Cell: Bacterial Behavior Reconsidered." In: *Frontiers in Microbiology* 6, article 264.
Lyon, Pamela (2017): "Environmental Complexity, Adaptability and Bacterial Cognition: Godfrey-Smith's Hypothesis under the Microscope." In: *Biology & Philosophy* 32, pp. 443–65.
Machamer, Peter/Darden, Lindsay/Craver, Carl. F. (2000): "Thinking about Mechanisms." In: *Philosophy of Science 67*, pp. 1–25.
Manneville, Paul (2006): "Rayleigh-Bénard Convection: Thirty Years of Experimental, Theoretical, and Modeling Work." In: Mutabazi, Innocent/Wesfreid, José Eduardo/Guyon, Etienne (Eds.): *Dynamics of Spatio-Temporal Cellular Structures*. New York, NY: Springer, pp. 41–65
Matthen, Mohan/Ariew, André (2002): "Two Ways of Thinking About Fitness and Natural Selection:" In: *Journal of Philosophy* 99, pp. 55–83.
Mayr, Ernst (1961): "Cause and Effect in Biology." In: *Science* 134, pp. 1501–6.
Mayr, Ernst (1982): *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Cambridge, MA: The Belknap Press of Harvard University Press.
McLaughlin, Peter (1990): *Kant's Critique of Teleology in Biological Explanation: Antinomy and Teleology*. Lewiston: E. Mellen Press.
Millikan, Ruth Garrett (1984): *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.
Moreno, Alvaro/Mossio, Matteo (2015): *Biological Autonomy*. Dordrecht: Springer.
Mossio, Mossio/Saborido, Christian/Moreno, Alvaro (2009): "An Organizational Account of Biological Functions." In: *The British Journal for the Philosophy of Science* 60, pp. 813–841.
Müller, G. B. (2017). "Why an extended evolutionary synthesis is necessary." In: *Interface Focus 7*, 20170015.
Neander, Karen (1991): "Functions as Selected Effects: The Conceptual Analyst's Defense." In: *Philosophy of Science* 58, pp. 168–84.
Nicoglou, Antonine (2015): "The evolution of phenotypic plasticity: Genealogy of a debate in genetics." In: *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *50*, pp. 67–76.
Okasha, Samir (2018): *Agents and Goals in Evolution*. Oxford, UK: Oxford University Press.
Paltridge, Garth W. (1979): "Climate and Thermodynamic Systems of Maximum Dissipation." In: *Nature* 279, pp. 630–31.
Pérez, Juana/Moraleda-Muñoz, Aurelio/Marcos-Torres, Francisco Javier/Muñoz-Dorado, José (2016): "Bacterial Predation: 75 Years and Counting!" In: *Environmental Microbiology* 18, pp. 766–79.
Pittendrigh, Colin S. (1958): "Adaptation, Natural Selection, and Behavior." In: Roe, Anne/Simpson, George Gaylord (Eds.): *Behavior and Evolution*. New Haven: Yale University Press, pp. 360–416.
Prigogine, Ilya (1947): *Étude thermodynamique des phénomènes irréversibles*. Liège: Desoer.
Salmon, W. C. (1989). *Four Decades of Scientific Explanation*. Pittsburgh, PA.: University of Pittsburgh Press.

Schlosser, Markus (2015): "Agency." In: Zalta, Edward N. (Ed.): *The Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/archives/fall2015/entries/agency/.
Schrödinger, Erwin. ([1946] 2013): *Statistical Thermodynamics.* New York, NY: Dover Publications.
Shani, Itay (2013): "Setting the Bar for Cognitive Agency: Or, How Minimally Autonomous Can an Autonomous Agent Be?" In: *New Ideas in Psychology* 31, pp. 151–65.
Simpson, George Gaylord (1944): *Tempo and mode in evolution.* New York, NY: Columbia University Press.
Simpson, George Gaylord (1953): *The Major Features of Evolution.* New York, NY: Columbia University Press.
Skewes, Joshua C./Hooker, Cliff A. (2009): "Bio-Agency and the Problem of Action." In: *Biology & Philosophy* 24, pp. 283–300.
Sterelny, Kim (2000): *The Evolution of Agency and Other Essays.* Cambridge, UK: Cambridge University Press.
Tversky, Amos/Kahneman, Daniel (1974): "Judgment under Uncertainty: Heuristics and Biases." In: *Science* 185, pp. 1124–31.
Varela, Francisco J. (1979): *Principles of Biological Autonomy.* Amsterdam: North Holland.
Waddington, Conrad Hal (1942): "Canalization of Development and the Inheritance of Acquired Characters." In: *Nature* 3811, pp. 563–65.
Walsh, Denis (2012): "Mechanism and Purpose: A Case for Natural Teleology." In: *Studies in History and Philosophy of Biological and Biomedical Sciences* 43, pp. 173–81.
Walsh, Denis (2015): *Organisms, Agency, and Evolution.* Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781316402719.
Walsh, Denis/Lewens, Tim/Ariew, André (2002): "The Trials of Life: Natural Selection and Random Drift." *Philosophy of Science* 69, pp. 429–46.
Weber, Andreas/Varela, Francisco J. (2002): "Life after Kant: Natural Purposes and the Autopoietic Foundations of Biological Individuality." In: *Phenomenology and the Cognitive Sciences* 1, pp. 97–125.
Wei, Yan, Xiaolin Wang, Jingfang Liu, Ilya Nememan, Amoolya H. Singh, Howie Weiss, and Bruce R. Levin (2011): "The Population Dynamics of Bacteria in Physically Structured Habitats and the Adaptive Virtue of Random Motility." In: *Proceedings of the National Academy of Sciences* 108, pp. 4047–52.
West-Eberhard, Mary Jane (1989): "Phenotypic Plasticity and the Origins of Diversity." *Annual Review of Ecology and Systematics* 2, pp. 249–78.
Woodward, James (2003): *Making Things Happen: A Theory of Causal Explanation.* Oxford, UK: Oxford University Press.
Wright, Larry (1973): "Functions." In: *The Philosophical Review* 82, pp. 139–68.
Wright, Larry (1976): *Teleological Explanations.* Berkeley, CA: University of California Press.