Naturalizing Biological Agency: Constitutive and Dynamical Strategies

Fermín C Fulda

Abstract:

The view that organisms are agents—and that organismal agency is fundamental to explaining biological phenomena—has become a central topic in the philosophy of biology (Walsh 2015; Moreno & Mossio 2015; Corning et al. 2023). Unlike standard causal-mechanical approaches, however, the concept of agency carries distinct teleological and normative implications that must be naturalized to be scientifically legitimate. But what exactly does naturalism require? And what counts as an adequate naturalization? I propose two desiderata: causal-location and explanatory indispensability, and compare two naturalistic accounts of agency—the organizational or constitutive account (OA) (Moreno & Mossio 2015) and the ecological or dynamical account (EA) (Walsh 2015). I argue that while OA satisfies causal-location at the cost of explanatory adequacy, EA achieves explanatory adequacy while remaining silent on causal-location. This leads to a dilemma between causal reductionism (OA) and teleological primitivism (EA), rooted in differing criteria for what naturalism requires. I distinguish two increasingly demanding grades of scientific naturalism: scientific emergentism and scientific essentialism, and argue that the dilemma arises from OA's commitment to the latter and EA's to the former. I conclude by showing how the emergentist criterion can resolve the dilemma by integrating OA and EA into a two-stage strategy that satisfies both desiderata.

> "[…] finalism is in the first place a dynamic character of a certain mode of existence, coincident with the freedom and identity of form in relation to matter, and only in the second place a fact of structure or physical organization […]."
>
> Hans Jonas, *The Phenomenon of Life*, 1966: 86

1. Introduction: Agency in Organismal Biology

The view that organisms are agents, and that organismal agency is fundamental to explaining biological phenomena such as development and evolution, has emerged as a central issue in the philosophy of biology (Walsh 2015; Moreno & Mossio 2015; Okasha 2018; Noble and Noble 2023; Corning et al. 2023). One motivation is the perceived limitations of a purely mechanistic-reductionist approach (e.g., Kirschner et al. 2000; Goldenfeld & Woese 2007; Bray 2012; Nicholson 2013; Keller 2014), and the growing recognition that explaining the adaptive character of organisms requires according a role to their purposiveness (e.g., Shapiro 2011; Walsh 2015; Noble 2017; Newman 2022; Sultan et al. 2022; Fulda 2023). In this section, I introduce the phenomenon of agency, demonstrate its pervasiveness in organismal biology, and argue for the need for a naturalistic account.

Agents are a class of goal-directed or purposive systems. Purposiveness is the capacity to attain and maintain a set of end-states or goals across a range of actual and counterfactual conditions, through different causal pathways, mechanisms, and from different initial conditions. The mark of this capacity is the persistence and plasticity of the system's dynamical profile or gross behavior (Russell 1945; Sommerhoff 1950; Nagel 1977; Walsh 2015; McShea 2012; Fulda 2017; Stovall 2024). Persistence is the ability to reach a goal robustly by maintaining a trajectory despite perturbations, while plasticity is the ability to do so by changing the trajectory to accommodate or adjust for perturbations. Two kinds of purposive systems can be distinguished. In artificial purposive systems, such as machines like guided missiles or thermostats, the goals are extrinsically determined by design and are hence derivative. In contrast, the goals of naturally occurring purposive systems are intrinsically determined by the systems themselves (Nicholson 2013). Such systems qualify as agents, and their autonomous capacity to pursue intrinsic goals constitutes their agency. While the most familiar example of agency is psychological—the capacity to act for reasons by mentally representing goals—organisms and their parts physiologically, developmentally, and behaviorally organize, maintain, and regulate themselves in intrinsically purposive ways to survive, reproduce, and realize their life cycle.

As a distinctive dynamical capacity, purposiveness enables a distinctive mode of explanation: teleology. Because goals can be reliably attained across a range of actual and counterfactual conditions, the capacities, activities, and parts of purposive systems can be explained not just by citing the conditions that cause or realize them, but by citing the goals they subserve. An activity or part occurs *in order to* attain or maintain a goal, or *because* the activity or part is necessary for attaining and maintaining the goal under the circumstances. Unlike causal explanation, teleology has normative implications. Given the goal and the circumstances, the system is required to produce the necessary means (Walsh 2015; Fulda 2017). Goals thus set standards or norms for evaluating the appropriateness of the means.

Consider a cheetah chasing a gazelle and the gazelle avoiding the cheetah. Had the gazelle jumped left, the cheetah would have turned left; had it gone right, the cheetah would have followed; had it run straight, the cheetah would have continued forward. These counterfactuals reveal a purposive dynamical profile. Catching and escaping are intrinsic goals that require particular responses rather than others, given the circumstances. As agents, these animals can succeed or fail in these pursuits. Regardless of how these behaviors are caused or realized, they can be explained teleologically by the fact that they are the necessary means to achieve these goals under the circumstances. Or consider an amoeba preying on paramecia. Though amoebae rely on chemotaxis and engulfment rather than running and tearing, a similar chasing-avoiding dynamic occurs, warranting a comparable teleological explanation. A hierarchy of avoidance behaviors in response to noxious or persistent stimuli has been demonstrated in the ciliate *Stentor roeseli*, based on the ability to switch between several different behaviors in a non-random order (Dexter et al. 2019). Despite different mechanisms—including psychological capacities in the mammalian case—the same type of predatory and avoidance goal explains these activities across metazoan and protozoan organisms.

The phenomenon of agency is not only manifested in behavioral activities or processes. Development exhibits a characteristically agential dynamical profile (Bertalanffy 1968; Walsh 2015; Sultan et al. 2022; Fulda 2023; Nadolski & Moczek 2023). Gene regulatory networks are robust to perturbations, and phenotypic plasticity enables ontogeny to adjust to environmental variation (Kitano 2004; West-Eberhard 2003). Organisms direct their genetic, epigenetic, and environmental resources to grow, differentiate, and ultimately realize their species-specific form, thus shaping their own developmental trajectories (Laland et al. 2015). As Newman (2014) puts it, "it is not that detailed mechanisms cannot be identified for these developmental processes, but that their details seem less important than the higher-level morphological 'attractors'" which represent "implicit purposes."

Physiological regulation and homeostasis—such as thermoregulation, glucose and blood pressure control, and hormonal balance—are paradigmatic cases of intrinsically purposive processes aimed at maintaining optimal internal function (Sommerhoff 1950; Nagel 1977; Turner 2017). As Kirschner and Gerhart (2005, 108–109) put it, organismal robustness "stems from a physiology that is adaptive." Consider neutrophils, a type of white blood cell that acts as a first responder to infection and inflammation. Like predatory agents such as cheetahs or amoebae, neutrophils chase pathogens such as bacteria, which in turn exhibit avoidance behavior. But unlike predators that seek nourishment, neutrophils engulf and neutralize pathogens to protect the host's health. More broadly, metabolic networks dynamically adjust metabolic flux to preserve viability under nutrient fluctuations (Ke et al., 2018), and metabolic plasticity enables cells to adapt to physiological and environmental stress (Spinelli et al., 2021). These processes exhibit the characteristic robust and flexible dynamical profile of purposiveness. Because these biochemical processes reliably produce outcomes necessary to maintain viability across varying conditions, they can also be explained by reference to the viability-preserving end-states they tend to attain and sustain.

Cellular agency is evident in sub-cellular molecular processes. DNA repair maintains genomic integrity in response to environmental (e.g., UV radiation) and internal (e.g., replication errors) perturbations. Genome reorganization has been described as "natural genetic engineering" (Shapiro 2011), highlighting its agential character. Similarly, protein folding occurs by selecting energetically favorable intermediate structures, or isomorphs, from a repertoire of stable options. This selection is not random but biased toward energetic efficiency, which is necessary for proper cellular function. This process can fail, thus preventing the cell from attaining and realizing its goals.[1]

[1] For agency at supra-organismal scales such as biofilms, colonies or swarms (Gordon 2023; Fulda 2023).

As these examples indicate, organismal biology is prima facie implicitly committed to agential explanations. However, the nature and role of agency remains a matter of debate. Despite this, there is broad consensus that any adequate account of biological agency must be naturalistic. Unlike the standard mechanistic approach, the concept of agency involves a commitment to the existence of goals, which play a teleological role in biological explanation. Historically, this commitment has been associated with supernatural views—such as backward causation or vitalism—and thus regarded as unscientific (Walsh 2008). Therefore, for agency to be a scientifically legitimate concept, the nature and role of goals must be naturalized. But what does naturalism require? What counts as an adequate naturalization? In this paper, I address these questions and show how they not only clarify the current dialectic in the debate over the nature and role of biological agency but also help to move the dialectic forward.

In Section 2, I argue that an adequate naturalization requires two distinct desiderata: causal-location and explanatory indispensability. In Section 3, I compare two naturalistic accounts of agency, the organizational or constitutive account (OA) (Moreno & Mossio 2015) and the ecological or dynamical account (EA) (Walsh 2015). I argue that while OA satisfies causal-location but not explanatory indispensability, EA is explanatorily adequate but silent about causal-location. This dialectic creates a dilemma between causal reductionism and teleological primitivism rooted in differing implicit criteria of what naturalism requires. In Section 4, I argue that scientific practice supports two criteria or "grades" of naturalism: scientific emergentism (G1) (Batterman 2005) and scientific essentialism (G2) (Ellis 2001). In Section 5, I argue that reductionism results from OA's commitment to G2, which motivates a strategy of naturalization by mechanistic descent, while primitivism arises from EA's adherence to G1 and its naturalization strategy by behavioral ascent. This emergentist strategy, I show, solves the dilemma by integrating both approaches into a single, two-stage naturalization: biological agency is an ecologically specified universality class that plays a primitive teleological role in a phenomenological theory of organismal dynamics. In turn, the underlying causal organization that realizes agential dynamics can be non-circularly specified in terms of closure of constraints as part of a theory of how agency is physically constituted.

This paper assumes two widely accepted premises in recent debates on biological agency. First, that biological agency need not presuppose mental capacities such as intentionality, rationality, or consciousness. The ordinary, pre-theoretical notion of agency may be psychological.

However, biological agency is a theoretical construct, not an analysis of the ordinary pre-theoretical notion (see Burge 2009; Fulda 2017). Second, that the problem of biological agency as intrinsic purposiveness is independent of, or at least orthogonal to, the analysis of evolutionary function in terms of natural selection. While each of these assumptions can be contested, they are not the subject of this paper. Instead, the paper focuses on two approaches to agency that endorse these assumptions. Since these approaches are realist, I do not consider purely heuristic approaches to agency (Lewens 2007; Okasha 2018; Desmond & Huneman 2020).

2. Two Desiderata for Naturalizing Agency

I argue that an adequate naturalistic account of agency should satisfy two conditions, causal-location and explanatory indispensability.

2.1. Causal-location

Naturalism is roughly the view that nature, understood as the causal order of the world, is all there is, and that scientific methods of investigation are the best way to understand the natural order (Papineau 2020). A core metaphysical commitment of scientific naturalism is physicalism, the view that the physical domain is fundamental because it is causally closed or complete or self-contained such that "no physical event has a cause outside the physical domain" (Kim 1993, 280). Closure implies that anything that makes a causal difference in the physical domain must itself be either identical to the physical or physically constituted or realized. The minimal criterion for something to count as natural is having a place within the causal structure of the physical domain. Agents, of course, make a difference in the physical domain to the extent that they act or do things and hence cause physical events to occur. So, to demonstrate that agency is natural, the conditions that realize or constitute agency in the physical domain must be specified in non-circular, non-agential terms.

However, the difference that agents make is unlike any other inhabitant of the physical domain and this calls for an explanation. What is unique about agents among physical systems is that agents don't just cause physical events—they cause them for a goal or purpose, or because doing so is necessary for attaining a goal given the circumstances. Similarly, agents are not just caused to act by the physical properties of their surroundings but by the significance or value that physical conditions have for attaining their goals. So, to render intelligible the difference that agents make as agents, teleological-normative concepts are necessary.

But the physical domain is intrinsically purposeless or valueless and hence normatively inert (Bedau 1992). This raises the question of what natural properties, if any, give meaning to these concepts and factual, empirical content to the explanations in which they figure. The worry is either that the difference that agency makes violates physicalism by involving a vitalist ghost in the molecular machine, or that the difference is merely heuristic and hence explanatorily spurious. To address these concerns and show that agents, as such, make a genuine difference within the causal structure of the physical domain, a naturalistic account must locate agency in that structure

Because the causal difference that agents make as agents is so unique, the explanation of this difference must be ineliminably teleological. If the difference that agents make could be exhaustively explained in terms of the sub-agential causal conditions that realize them in the physical domain, agency would be theoretically and hence explanatorily superfluous. Therefore, it is not enough to locate agency in the causal order—it is also necessary to show that agency plays an indispensable teleological role in explaining causal order. But what exactly is a teleological explanation? How does it differ from causal-mechanical explanation?

2.2. Explanatory-Indispensability

Scientific explanation can be conceived as a relation of modal (counterfactual) dependence between the explanans and the explanandum that enables us to answer questions of the form what-if-things-had-been-different? (Woodward 2000; Strevens 2008; Walsh 2015). In causal-mechanical explanations of the form 'e occurred because e was caused by c', where the effect 'e' is the explanandum and the cause 'c' is the explanans, the occurrence of e causally depends on the occurrence of c. The modal profile of this dependence can be specified in terms of the following counterfactuals: If c causes e, then if c were not to happen, e wouldn't; and if c were to happen, e would. That this dependence holds can be empirically shown by intervention (Woodward 2000): Changing the value 'c' while keeping background conditions constant systematically changes the value 'e'. The explanatory relevance of this causal dependence is made intelligible by describing the mechanism—the set of parts, interactions, and organization—that brings about or produces the effect (e.g., Machamer et al. 2000). This description explains why e occurred rather than not by citing the antecedent occurrence of c.

Teleology conforms to this framework while involving a different kind of dependence, contrast class, and (normative) content (Walsh 2012; 2015). In teleological explanations of the form ‘S did/does ϕ in order to ψ (or: S’s ϕ-ing was directed at ψ-ing) in C’, where ‘S’ is a goal-directed or purposive system, ‘ϕ’ is the means or explanandum, ‘ψ’ is the goal or explanans, and ‘C’ are the circumstances, the occurrence of the means ϕ depends on the goal ψ in C. Following Aristotle, this counterfactual dependence between means and goals is not a causal relation but a relation of “hypothetical necessity” (Cooper 1987; Walsh 2008): ϕ is hypothetically necessary for ψ just in case ψ is a goal and ϕ is necessary for (or no worse than anything else/good enough for) the attainment of ψ under the circumstances C. The modal profile of this dependence can be specified in terms of the following counterfactuals (Sehon 1994; Walsh 2012): If the fulfillment of ψ had required ϕ^*-ing rather than ϕ-ing, then (ceteris paribus) S would have ϕ^*-d rather than ϕ-ing. And if the goal had been ψ^* rather than ψ, then (ceteris paribus) S would have ϕ^*-d. These counterfactuals reflect the persistence (robustness) and plasticity (flexibility) characteristic of S as a purposive system. The idea is that S would typically do whatever is (hypothetically) necessary to attain its goal ψ across a range of conditions C. That this dependence holds can also be empirically shown by intervention: Changing the value ‘C’ (circumstances) while keeping the value of ‘ψ’ (goal) constant or vice versa systematically changes the value ‘ϕ’. The explanatory relevance of this dependence is made intelligible by describing the way the occurrence of the means is conducive to the attainment of the goal given the circumstances. This description explains why ϕ occurred rather than ϕ^* by citing the goal ψ and circumstances C.

Teleological explanation has distinctive normative implications. If S has goal ψ and ϕ–ing is hypothetically necessary to attain ψ, then (ceteris paribus) S is required to ϕ. On this basis, we can evaluate whether ϕ–ing is appropriate or inappropriate for attaining ψ under the relevant circumstances (Walsh 2015, 201). As such, goals set standards or norms for evaluating the appropriateness of the action given circumstances.

Thus, while mechanisms causally produce their effects, goals normatively require their means. This makes teleology prima facie a distinctive non-causal form of explanation. However, it doesn’t follow that causal relations are not part of teleology or that the same phenomenon cannot be the explanandum of both kinds of explanations. Causal relations involved in teleology include the fact that ϕ, the means, is caused by S, the agent, and the fact that if ϕ is successful in attaining ψ, the goal, then ϕ causes ψ. However, none of these causal relations teleologically explains why—

that is, for what goal or purpose—S did ϕ rather than ϕ^* or nothing at all. For that, we need to cite the goal ψ and describe the way the occurrence of ϕ hypothetically depends on, or is required by, ψ—and not the way ϕ causally depends on S. After all, due to the persistence and plasticity of S, the modal profile of hypothetical dependency implies that had ϕ not happened, some other means ϕ^* would have occurred to attain ψ across a range of changes in the values of ψ and C. So, as Walsh (2008; 2012; 2015) has emphasized, the means is subject to two different yet complementary explanations—causal and teleological—on account of the fact that two different but complementary relations of counterfactual dependence are instantiated in purposive systems.

In summary, since agents make a difference in the causal structure of the physical domain, an adequate naturalistic account must locate agency in this structure. The conditions that realize agency in the physical domain exhibit the characteristic modal profile of causal relations and hence must be specified in non-agential, strictly causal-mechanical terms. But the unique difference that agents make as agents in the physical domain has the characteristic modal profile of means-ends relations (hypothetical dependence). Therefore, this difference cannot be captured in purely causal-mechanical terms and must instead be specified in irreducibly teleological-normative terms. Thus, an adequate naturalistic account of agency must satisfy not just causal-location but also demonstrate that the teleological-normative concepts implicated in agency are explanatorily indispensable.

3. Two Accounts of Agency: The Constitutive and the Dynamical

Let us now consider two naturalistic accounts of agency and see whether and how they meet these desiderata.

3.1. The Organizational or Constitutive Approach

According to the organizational approach (OA), agency is the capacity of an 'autonomous' system to interact with the environment in a way that maintains its internal organization (Varela 1979; Maturana and Varela 1980; Christensen and Hooker 2002; Di Paolo 2005; Thompson 2007; Barandiaran et al. 2009; Arnellos and Moreno 2015; Moreno and Mossio 2015). Autonomous systems belong to the generic class of thermodynamically open, far-from-equilibrium, self-organizing, self-maintaining systems. What distinguishes autonomous systems is the specific causal regime of their internal organization which involves a constitutive dimension, ‘closure of constraints’, and an interactive dimension, ‘agency’ (Moreno and Mossio 2015). ‘Closure of constraints’ consists of a recursive network of interdependent components (constraints) which together produce and maintain themselves as a functionally integrated whole (Moreno and Mossio 2015; Montévil and Mossio 2015). The paradigmatic example is a metabolic system such as a single cell, constantly exchanging energy and matter with the external environment to dynamically produce and maintain its organization far from equilibrium. In turn, the interactive dimension is a subset of these constitutive constraints characterized by the function of adaptively regulating the environmental interactions of the system. The paradigmatic example is bacterial chemotaxis, where the flagellum is the agential contribution to realizing metabolic closure by constraining movement in the direction of the source of nutrition.

According to OA, “the maintenance of the whole organization can be taken as the naturalized goal of agential functions, and its conditions of existence are the norm of their activity” (Moreno and Mossio 2015, 93). The goal is intrinsic to the system because it is determined by its internal physical constitution. This determination is a part-whole relation of circular causation: The goal is the effect of the activities of the system and this effect, in turn, is the cause of these activities. “Biological systems are teleological because the effects of their own activity contribute to establish and maintain their own conditions of existence” (Mossio & Bich 2017). Teleological explanations are thus causal explanations of a system’s self-maintenance. These causal explanations are supposed to have normative implications: Because the states or effects with which goals are identified *must* occur for the system to exist, the system *ought* to cause them. The circular causal relation between the system and its own maintenance thus specifies the norms that the system must conform to in order to exist or persist. Activities that fail to contribute to the maintenance of the system and hence fail to conform to these norms are thus not doing what they are *supposed* to do.

By defining agency physically as an interactive constraint on the constitutive organization of the system, OA locates agency in the thermodynamic structure of the physical domain. The central idea is that agency is a consequence of the thermodynamic openness of organizationally closed systems. OA thus satisfies causal-location by showing that commitment to goals and norms as causal contributions to self-maintenance is consistent with physicalism. Since an agent’s goals are local, physical, causally efficacious properties, they play a genuine role in the causal explanation of the behavior of the system. Moreover, this definition offers a criterion of demarcation in terms of the physical constitution of the system that allows us to distinguish between genuine agents that have goals intrinsically, such as living cells. It also distinguishes these from mere self-organizing but non-agential systems such as hurricanes or flames, and extrinsically goal-directed systems such as artifacts. OA is thus a parsimonious, realist account of natural agency in which no new entities are introduced over and above the physical constitution and causal dynamics of the system.

But is OA explanatorily adequate? One problem is that OA misconstrues the form of teleological explanations. Even assuming, for the sake of argument, that being necessary for self-maintenance is a genuinely normative relation, it is not the kind of normative relation that holds between a goal and its means. The former is a relation of *causal* dependence, but the latter is a relation of *hypothetical* dependence. These relations have different modal profiles that support different modes of explanation—causal and teleological, respectively. Being causally necessary for self-maintenance is thus the wrong kind of dependence for teleology. In fact, due to the robustness of the system, the particular means may be causally sufficient but not necessary for the goal. This provides no information about what the system would have done in counterfactual circumstances, and hence no sense in which the response is normatively required for attaining the goal given the circumstances or for evaluating the response as appropriate.

Another problem is the extensional inadequacy of OA's definition of goals. There are many things that a system is required to do—for example, to attain some biological goal—that are not preconditions for the system's continued persistence. Purposive systems can typically attain states that constitute malfunction without thereby ceasing to exist. Rather, the system continues to exist but does so at a less-than-optimal level (Moosavi 2019). However, OA makes teleological normativity an all-or-nothing matter: failure does not imply non-existence. An animal, for example, can have the goal of eating an insect, and its thermoregulatory system may be directed at the goal of slightly lowering the animal's temperature. The animal won't die if it fails to achieve either of these goals. So, by identifying goals with self-maintaining effects, the definition is both too inclusive and too restrictive.

To the extent that OA physically defines agency in strictly causal terms and takes teleology to be an instance of causal explanation, OA is a reductionist approach. As such, it makes a fundamental contribution to locating agency in the causal structure of the physical domain. But it does so at the expense of vindicating the distinctive teleological form and normative content of agential explanations. The phenomenon of agency, it seems, cannot fit within a purely physical, causal-mechanical framework. OA may respond that even if it makes teleology an instance of causal explanation, it need not make teleology an instance of mechanical explanation—and this, they may argue, is enough to fulfill the anti-reductionist aspirations that the concept of agency is called to play in biology. After all, not all desiderata are guaranteed to be met by a scientific analysis, and the demand for explanatory indispensability in the strict teleological-normative sense is arguably one of them.

3.2.The Ecological or The Dynamical Approach

An alternative approach holds that agency is the gross dynamical capacity of an ecologically embedded system to bias its repertoire in response to its affordances in pursuit of its goals (Walsh 2015; 2018; Fulda 2017; 2023; Sultan et al. 2022; Moczek and Nadolski 2023; Walsh & Rupik 2023). Recent studies in ecological evolutionary developmental biology (Gilbert and Epel 2015) show that, far from being a mere constraint on selection, development enables adaptation by facilitating non-random phenotypic variation (Gerhart and Kirschner 2007; Salazar-Ciudad 2021; Pfennig et al. 2021). The adaptive bias that development introduces to the generation of biological form is the effect of ecologically responsive phenotypic capacities of individual organisms such as 'developmental bias' (Uller et al. 2018), 'phenotypic plasticity' (Pigliucci 2001; West-Eberhard 2003), and 'niche construction' (Odling-Smee et al. 2003). Through these capacities, organisms direct their ontogeny towards adaptive, beneficial outcomes by mediating (biasing) the effects of their genomic, epigenomic, and environmental conditions (Fulda 2023; Sultan et al. 2023; Nadolski and Moczek 2023). It is this organism-centric theory of development that EA argues is committed to agency.

On this view, a goal is a state of affairs G only if there is a system S that tends to attain and maintain G in a robust (persistent), adaptive (plastic) way across a range of conditions by biasing its repertoire R in response to the affordances that S encounters and co-constructs. So, a goal is just an end-state that tends to be achieved in a certain way. A repertoire R, in turn, is a biased range of potential responses r {r1, r2..., rn} that enables S to realize its goals G in response to its affordances. An affordance x is a property of an organism-environment system considered a single, coupled dynamical system that impedes or promotes the deployment of a system's repertoire R in pursuit of its goals G. The concept is meant to capture the constitutive reciprocity or interdependence between organism and environment (Gibson 1979; Lewontin 2001; Walsh 2012). This triad of concepts forms an inter-defined theoretical cluster that constitutes an agential dynamics, a method for explaining the behavior of agents as such by citing ecological facts about the organism-environment system. According to this method, a system S would do what is (hypothetically) required to attain a goal G by biasing its repertoire R in response to its affordances x across a range of actual and counterfactual values for these variables. The agential dynamical profile of S is thus characterized by a distinctive hypothetically invariant modal profile.

Consider a bacterium swimming up a glucose gradient. A causal-mechanical explanation of this behavior cites sub-organismal mechanisms and pathways—it explains how the activity is produced and why it occurred rather than not. However, it does not account for the fact that, had this particular pathway failed or environmental conditions differed, some alternative configuration from the organism's behavioral repertoire would have been recruited. Explaining such (hypothetical) invariance requires reference to the agential dynamics that shapes the organism's ecological phase space by constraining the range of possible trajectories. Moreover, the adaptive value of the outcome—whether the bacterium successfully reaches the glucose—is incidental to the mechanical explanation. The activity occurs because it is the effect of antecedent causes, not because it is conducive to nourishment. By contrast, an agential explanation accounts for the behavior by reference to its goal: the bacterium swims up the glucose gradient in order to attain glucose. This is because glucose affords nourishment, the bacterium has the metabolic need or goal of acquiring nourishment, and the presence of glucose biases its chemotactic repertoire toward a specific trajectory—combining straight runs and random tumbles—that leads to the glucose. In this sense, the chemotactic trajectory occurs not merely because it is caused, but because causing this response is hypothetically required to attain the goal under given ecological conditions. Ceteris paribus, if the goal or the affordances had been different, a different chemotactic pattern would have been selected from the repertoire. This also implies that the particular response that was produced from the repertoire can be normatively evaluated as appropriate or inappropriate with respect to the goal and the affordances. Therefore, unlike in the causal-mechanical explanation, the success of the activity is not incidental to the explanation—it is essential. Agential dynamics thus provides irreducible ecological information about the hypothetically invariant patterns of bacterium–glucose interaction that are necessary to explain the adaptive character of chemotaxis.

The teleological force of agential dynamics has distinctive ontological commitments. The identity of a glucose molecule as a chemical compound is, of course, determined by its intrinsic microstructural physical properties. However, in the presence of bacteria with a certain goal-biased metabolic and behavioral repertoire, glucose is transduced into a metabolite and hence into a positive affordance for attaining nutrition. The metabolic significance of glucose is thus an ecological property jointly constituted by the metabolic repertoire of the bacterium and the microstructural properties of the molecule. As a metabolic affordance, glucose is an emergent, qualitatively distinct property of the whole bacterium-environment system. So, even if as a chemical compound glucose is physically and hence causally individuated, as a nutrient it is ecologically and hence normatively individuated. Similarly, the identity of the bacterium as a physical system can be specified irrespective of its surroundings and causally explained in terms of the mechanisms, pathways, and internal molecular organization that constitutes it as a differentiated physical entity. However, it is only when we consider the bacterium as embedded in its chemical 'affordance landscape' (Walsh 2021) that the identity of the bacterium as an organism pursuing a way of life characteristic of its kind emerges. The goals of the bacterium are intrinsic not because they are physically internal, but because they are immanent to its gross ecological dynamics.

According to EA then, agency is not a local, physical property of the internal organization of the bacterium that explains by causing chemotaxis. Rather, it is an emergent, dynamical, ecological property of the whole bacterium-glucose system that explains by normatively requiring chemotaxis. EA thus vindicates the characteristic non-causal, teleological-normative role that an agent's goals play in explanation. Explanatory indispensability is accomplished because the method of agential dynamics is committed to an ontology of emergent, ecological, normatively individuated properties of the whole organism-environment system, including goals, repertoire, and affordances, governed by a relation of hypothetical invariance. Indeed, the hypothetically invariant relation that governs agency as a dynamical mode and teleology as mode of explanation is built into the (circular) ecological tripartite definition of agency as a primitive theoretical principle. Explanatory adequacy comes at the expense of ontological parsimony on indispensability grounds.

EA may be explanatorily adequate, but it offers no positive account of the place of agency in the causal structure of the physical domain. To focus on the ecological regularities at the organismal scale, the physical constitution of the system is treated as a black box. Defenders of EA have been content to point out how much we already know about the causal-mechanical conditions underpinning purposive behavior to be confident that agential dynamics does not violate physicalism (Walsh 2008; 2015; 2018; Fulda 2017; 2023). But these conditions, they insist, are not part of the theoretical account of the nature and role of agency. Agency is a primitive, circularly defined ecological concept in a theory of organismal dynamics. As such, although consistent with the requirement that instances of agency must be wholly physically realized, the definition does not specify causal conditions for applying the concept. There is thus no strict demarcation between systems that can and cannot be explained using agential dynamics. Although there are clear-cut cases, there are bound to be borderline cases.

3.3. A Dilemma between Causal Reductionism and Teleological Primitivism

Let's take stock of the dialectic. OA and EA both seek to naturalize agency, and both take agency to be an interactive property of organisms understood as self-producing, self-organizing, self-maintaining and self-regulating systems. But while OA defines this interactive property reductively in terms of the internal causal organization of an organism's physical constitution, EA defines it ecologically in terms of the organism's gross purposive capacity to respond to conditions as affordances. OA thus gives a parsimonious account of the place of agency in the thermodynamic structure of the physical domain. However, it does so at the expense of failing to preserve the teleological form and content of agential explanations. Conversely, EA gives an explanatorily adequate account of agency by committing itself to a set of primitive ecological properties that play an irreducible teleological theoretical role. However, it necessarily says nothing about the causal-mechanical conditions that realize these emergent properties. So, despite their merits, neither of these accounts offers a complete naturalization of agency.

This dialectic leads into a dilemma between the causal reductionism of OA and the teleological primitivism of EA: Either agency is part of the causal structure of the physical domain but an agent's goals do not play a distinctive teleological-normative role in explaining biological phenomena, or an agent's goals play a distinctive teleological-normative role in explaining biological phenomena but there is nothing specific to be said about the place of agency in the causal structure of the physical domain. The fact that both accounts take themselves to be naturalizing agency, indicates that the dilemma is predicated on different views about what naturalism requires and what the proper method of naturalization is or should be. Making these assumptions explicit will allow us to diagnose the dilemma and to determine whether and how it can be solved allowing for the possibility of an integrated, complete naturalistic account of agency.

4. Two Grades of Naturalistic Involvement

I argue that scientific practice indicates two criteria or grades of naturalistic involvement, the second more demanding than the first one.

4.1. Grade 1 Naturalism: *Scientific Emergentism*

According to the first criterion (G1), X is natural iff X has a place in the causal structure of the physical domain, and the concept of an X plays an indispensable explanatory role in a scientific theory. The first condition is just the minimal metaphysical commitment of naturalism to physicalism that motivates causal-location. The role of causal-location is to ensure that agency incurs in no supernatural commitment that violates causal closure. As such, this condition says nothing about X's theoretical identity and role. However, many things are said to be natural by virtue of playing an indispensable role in a scientific theory irrespective of their physical constitution and hence irrespective of their causal-location (Stich 1994; Tye 1994; Burge 2010). To capture this, the second condition adds a theoretical, explanatory condition to physicalism that expresses the basic commitment of naturalism to science. For the concept of an X to have a place in a scientific theory is for X to play an indispensable role in the explanation of a distinctive set of empirically observable and counterfactual-supporting regularities. So-called 'universal' phenomena in the physics of complex systems such as phase transitions and critical phenomena in condensed matter physics and statistical mechanics are clear instances of this criterion. Call this criterion scientific emergentism.[2]

[2] I adopt the widely accepted view that scientific explanation involves both epistemic and ontic components (Cartwright, 1983; Salmon, 1984; Machamer et al., 2000; Psillos, 2002; Woodward, 2003; Strevens, 2008). Thus, the explanatory criterion (G1) is not purely epistemic; it carries an ontological commitment. In particular, if a concept or property is indispensable to our best explanations, this warrants commitment to its existence. Given this, the notion of emergence employed here is one of weak emergence: an emergent property or entity, X, is not identical to its physical realizer or constitution, yet it must be physically realized or constituted, thereby respecting physicalism. By contrast, strong emergence rejects this constraint and consequently violates physicalism (O'Connor, 2021; Wilson, 2021). For alternative accounts of emergence, see Mitchell (2003) and Gillett (2016). I am grateful to an anonymous reviewer for prompting this clarification.

Properties such as viscosity, superconductivity or ferromagnetism constitute universality classes, that is, sets of physical systems that exhibit the same or similar macroscopic behavior near a critical point independently of the microscopic details of each individual system. 'Universality' refers to the generic dynamical properties of macroscopic behaviors that do not depend on the details of the physical constitution of the system that realize them (Batterman 2000; 2006; 2018; 2021; Morrison 2012; 2014; Morrison et al. 2015; Rice 2019). The reason is that the large-scale macroscopic properties or behaviors of the system have a distinctive dynamical profile—undergo a different set of changes—from the lower-scale microphysical properties or conditions that physically constitute or realize them, making them multiply realizable. In other words, the macroscopic behaviors are robustly invariant across changes in the underlying micro-realizers. Phenomenological (macroscopic) models that represent the gross universal dynamics of the system thus use a distinctive set of primitive macroscopic concepts that cannot be reductively defined in terms of their microphysical realizers. Rather, the concept is inter-theoretically defined along with other primitive concepts at the higher level of organization. As Batterman (2005) puts it,

> "the bulk or gross properties of fluids can be studied as a universality class irrespective of the details of the microphysical configurations that realize them."

Consider viscosity, the gross behavioral disposition of a fluid to resist gradual deformation (flow) under shear stress. Viscosity is located in the molecular structure of the physical world: it is the effect of the internal friction of the individual molecules that physically constitute the fluid on account of their shape, momentum, and charge. So, viscosity—a large-scale behavioral property of fluid—has a lower-scale molecular hence structural realizer. However, despite the discrete ontology of a fluid's microphysical constitution, models of fluid dynamics that use the concept of viscosity represent the fluid as a continuum. This idealization indicates that viscous behavior, and fluid behavior more generally, is largely insensitive to the details of the microphysical constitution that realizes it. After all, fluids with very different microphysical constitutions such as honey, water, or milk, can instantiate the same large-scale viscous behavior. The large-scale universal features that constitute the gross dynamics of viscous behavior are thus robustly invariant, displaying a distinct modal profile from their micro-constitution (Batterman 2005; Rice 2019). As a universal mode of behavior, the underlying microstructure of a viscous fluid thus makes no difference, for the most part, to the prediction and explanation of the gross dynamics of the fluid under shear stress. The difference that viscosity makes in models of fluid dynamics calls for a distinctive set of concepts that have no microscopic counterpart.

Because of the dynamical and modal autonomy of the large-scale universal behavior of fluid, there is no set of individual molecules the number of which is necessary and sufficient for the concept of viscosity to apply to a fluid. The concept is rather interdefined with other theoretical primitive concepts, such as density and surface tension, at the same gross dynamical scale of the continuum hypothesis. There is thus no strict demarcation between genuine fluids in microscopic terms. Rather, there are paradigmatic and borderline cases that can only be determined on irreducible behavioral grounds. Viscosity is nonetheless as natural as can be on account of the concept playing an indispensable macroscopic role in fluid dynamics. It is then the gross dynamical theoretical role of the concept, rather than its underlying microstructural realization, that specifies the nature or identity or principle of individuation of viscosity: Viscosity is what it is—a phenomenologically empirically accessible mode of macroscopic behavior—rather than something else, a microscopic molecular configuration. As Goldenfeld and Kadanoff (1999: 87–88) put it,

> "In fluid dynamics the large-scale structure is independent of the detailed description of the motion of the small scales. [...] To get these gross features, one should most often use a more phenomenological and aggregated description, aimed specifically at the higher level [...] by trying to separate universal scaling features from specific features."

Consider superconductivity, the disposition of a substance to carry electricity without resistance below a critical temperature. Superconductivity has a place in the quantum order of the world, as it is causally realized by the pairing of the electrons that physically constitute the superconductor. The microscopic theory of superconductivity is called the "Bardeen-Cooper-Schrieffer" (BCS) theory. However, the concept of superconductivity plays an indispensable macroscopic role in the "Ginzburg-Landau theory" in condensed matter physics. This theory gives a large-scale phenomenological description of superconductivity as a universal property that is largely insensitive to the conditions that realize it in terms of the BCS theory (Morrison 2015). As Goldenfeld and Woese (2011, 378–379) put it:

> "There is nothing fundamental about the atoms or molecules. [...] The phenomenon of superconductivity as a process is captured by the universal, symmetry-based Ginzburg-Landau theory, but that process can have many different realizations or instantiations from matter at a variety of energy and length scales."

4.2. Grade 2-Naturalism: *Scientific Essentialism*

According to the more demanding grade, X is natural iff all Xs have a place in the causal structure of the physical domain; the concept of an X plays an indispensable explanatory role in a scientific theory (G1); and X is a natural kind individuated by its causal realizer (G2). G2 adds two essentialist conditions to G1. First, X is individuated in terms of the necessary and sufficient conditions that realize X in the causal structure of the physical domain. Causal-location implies causal-individuation. Second, this implies that the theoretical role of the concept of an X can be specified and hence defined non-circularly by the causal conditions under which the concept applies. Together these essentialist conditions imply that to say that X is natural is to say that X can be defined in non-circular strictly causal terms by the set of necessary and sufficient conditions that causally realize X.

This criterion is familiar from the microstructural essentialist approach to natural kinds (Kripke 1980; Putnam 1976; Ellis 2000; Bird 2007). On this view, having a common intrinsic structure is necessary and sufficient for being a member of the kind X. This structure determines the identity and existence conditions of X, explains its general features, and strictly demarcates X from other kinds. The paradigmatic examples are chemical elements. A large collection of molecules is properly classified as a fluid on account of its invariant collective universal dynamics irrespective of its multiple microphysical constitutions. But a lump of metal is properly classified as gold on account of its microstructural composition (atomic number), which explains the behavior of the lump including its malleability, electrical and thermal conductivity, density, etc. The microstructure demarcates instances of gold from other chemical elements. So, whereas the identity of viscosity is not only manifested but constituted by its behavioral phenomenology, the essence of gold is manifested by its behavioral phenomenology but is not constituted by it—it is a microstructural property, not a mode of behavior.

Consider the case of oxygen. Oxygen satisfies G1 by virtue of having a distinctive place in the sub-atomic structure of the physical domain. The concept of oxygen, in turn, satisfies G2 by virtue of playing an indispensable role in chemistry, the science of the structure and transformation of matter. Oxygen is a kind of atom and hence a constituent of matter. As such, it is invariant across changes in the structure of matter. But the role that oxygen plays in chemistry is strictly specified by its microphysical structural constitution as per G1—its atomic realizer—so oxygen satisfies G2. Having eight protons determines the identity of an atom of O in the strict sense of being necessary and sufficient for that atom to be an instance of that kind and hence provides a criterion of demarcation. Indeed, the atomic number specifies necessary and sufficient atomic conditions for a substance to be a member of the kind such that the atomic conditions for the application of the concept 'oxygen' exhaust its content and role in scientific generalizations. Furthermore, the atomic number also determines nuclear charge, the most invariant property across chemical changes, for it explains the direction of chemical change (Hendry 2008). So, the microstructural (sub-atomic) conditions for applying the (atomic) concept exhaust its content and non-circularly specify its theoretical role.

In summary, although specifying the conditions that realize agency in the physical domain is a necessary condition for an adequate naturalization, it is not sufficient. Scientific practice indicates two importantly different ways of understanding the form and force of this specification: the standard essentialist way (G2), where causal-location takes the form of a strict causal definition and has the force of a principle of individuation; and the emergentist way (G1), where causal-location takes the form of a causal or realization explanation, not definition, allowing emergent properties to be individuated by their own gross dynamical profile as specified by their theoretical role, irrespective of their realizer.

5. An Integrated Account of Natural Agency

I argue that these criteria allow us to diagnose and solve the dilemma.

5.1. Causal Reductionism, Scientific Essentialism, and Mechanistic Descent

We can now see that the reductionist horn of the dilemma is the result of OA's commitment to the essentialist G2 criterion of naturalism. This commitment is manifested in the assumption that the physical constitution of the system not only *locates* agency in the causal structure of the physical domain by specifying the causal organization that realizes agential behavior, but it also *defines* agency in non-circular causal-mechanical terms. Since 'agency' is essentially a teleological-normative concept, defining agency causally amounts to a reductionist definition. Having defined agency causally, an agent's goals can only play a causal role in the explanation of the behavior of the system, thus losing the characteristic teleological-normative form and content of agential explanations. The explanatory inadequacy of OA and hence the reductionist horn of the dilemma is thus predicated on OA's commitment to the G2 criterion of naturalism.

This commitment is explicit in its method of naturalization by what we can call 'mechanistic descent.' According to this strategy, a scientific account of agency must not only explain but also define the behavioral phenomenology in terms of the internal causal-mechanical organization of the system. Barandiaran et al (2009, 6) say:

> "From a descriptive [phenomenological] standpoint one could … use of these conditions [individuality, interactive asymmetry and normativity] to evaluate whether a given system is an agent or not and to test some available models. But a proper definition of agency… must specify what is the generic and minimal type of system or mechanism that is capable to generate, by itself, the properties that meet these conditions. […] [T]his requires that we look inside, that we explain these features in terms of how the system is organized and organizes its interactions with the environment. As Rohde and Stewart (2008) argue, the ascription of this kind of features on solely behavioral grounds (if possible at all) stands on much weaker feet than those grounded in scientific study of the underlying mechanisms involving the organization of the agent."

Similarly, Bich et al. (2016, 242) claim that

> "Substantial advances can be made […] if we […] try shifting from a mere phenomenology of compensatory behaviors (focused on generic responsive capacities), towards a more precise, *organisational* account of the distinctive features of the mechanisms responsible for those different behaviors. […] the problem lies in the tendency to focus on the *effects*—i.e. the systems having adjusted

itself in such a way to counter the perturbation—rather than on the *nature* of compensatory responses—i.e. *how* the response is achieved […] *how* those adaptive skills are actually implemented in each of the cases." (Emphasis in the original)

So, for OA's strategy of naturalization by mechanistic descent, the phenomenological-behavioral level of description is at worst not properly scientific, and at best merely heuristic or 'ascriptional.' Either way, it cannot be the level at which agency is naturalistically defined. Only the underlying mechanical level can. This reductionist assumption is clearly stated by Nicholson (2013, 674), who assigns a mere epistemic role to the behavioral level:

> "Although the behavioral pattern of a system affords good evidence of its purposiveness, the purposiveness of a system cannot be explained in terms of observable behavior—much less be defined in terms of inputs and outputs whilst black-boxing the system's internal organization causally responsible for it. […] The purposiveness of a system does not depend on its behavioral response patterns but on the internal organizational regime causally responsible for them."

In short, the explanatory inadequacy of OA and hence the reductionist horn of the dilemma is a consequence of OA's commitment to the G2 criterion which thereby motivate a G2 strategy of naturalization by mechanistic descent.

5.2. Teleological Primitivism, Scientific Emergentism, and Behavioral Ascent

EA asserts precisely what OA denies, namely, that the behavioral properties of agency manifested in the phenomenology are enough to naturalize it, provided each agential system has a causal realizer. EA is thus committed to the G1 emergentist criterion, which is explicit in EA's method of naturalization by what we might call 'behavioral ascent'. On this view, Nicholson is surely right that behavioral patterns constitute evidence of purposiveness. 'Depend', however, is ambiguous between a causal and an identity claim. No doubt, these patterns *causally* depend on, and are physically constituted by, the internal causal organization. But unless an essentialist criterion and the concomitant strategy of mechanistic descent is assumed (G2), it doesn't follow that the purposive identity and teleological role of these patterns depends on the internal causal organization. This identity and role could be irreducibly behavioral, as universal properties such as viscosity, superconductivity or ferromagnetism are (G1).

It also doesn't follow that the explanation of why these purposive patterns or behavioral properties occur cannot cite further patterns or behavioral properties. As Morrison (2015: 105) puts it, "we need not appeal to the micro phenomena to explain the macro processes." Models of fluid behavior often explain viscosity by citing other gross dynamical properties of the fluid such as density or pressure irrespective of how these dynamical properties are structurally realized. According to EA, just as flowing behavior is invariant across different realizations, so is acting or responding purposively. These emergent patterns undergo their own distinctive set of changes and hence have their own dynamical and modal profile. Therefore, they are explanatorily indispensable as such. To deny this on the essentialist assumption that the identity and explanatory role of system-level, phenomenologically accessible behavioral properties *must* be determined by the underlying conditions that realize them as per G2 is to contradict the standard practice of behavioral ascent in the physics of complexity. Given this emergentist alternative, the epistemic role of behavior in detecting purposiveness (or viscosity) is compatible with this behavior also playing an ontological role in individuating purposiveness (or viscosity) as such and defining its theoretical role. This, in turn, is compatible with the internal causal organization of the system playing the role of explaining—but not defining—*how* (rather than why) the behavior is caused or realized.

Pace OA, the scientific study of the dynamics of fluids does not require that "we looked inside" the system into its physical constitution, for the most part, nor does the naturalistic status of gross behavioral properties such as viscosity or superconductivity depend on being defined in terms of the fluid's physical constitution. Instead, the scientific study of the behavior of fluids proceeds by behavioral ascent from the properties of the individual molecular components that physically constitute the fluid, to the properties of the whole collection of components that characterized the fluid's gross pattern of behavior. As per G1 then, these properties are part of the causal molecular structure of the microphysical domain, but their microscopic location or realizer does not figure in the specification of the role that the macroscopic properties play in phenomenological models of fluid dynamics. Rather, those macroscopic properties are primitive concepts in these models.

This and other cases of phenomenological models of the large-scale behavior of complex system that are insensitive to and hence autonomous from its underlying micro realizers thus offers a scientific model and justification for the emergentist strategy of naturalization by behavioral ascent of EA. Just as viscosity is a universal behavioral response pattern of a fluid to shear stress that plays an indispensable macroscopic causal role irrespective of its microphysical constitution, purposive agency is a 'universal' biological response pattern to positive or negative affordances that plays an indispensable teleological-normative role irrespective of its internal causal-organization. Generic behaviors such as swimming, eating, hiding, escaping, chasing, etc., or physiological activities such as respiration, digestion etc., are all phenomenologically identifiable by behavioral ascent and may be considered to belong to the same agential universality class despite the fact that they are realized across a wide range of physical constitutions. It is therefore scientifically legitimate for the EA to abstract away the details of causal implementation by treating the internal physical constitution of the system as a black box, and focus instead on the behavioral patterns by introducing primitive concepts apt to capture the dynamical (and modal) profile of these patterns.[3]

Agential dynamics is a method for explaining the behavior of agents, just as fluid dynamics is a method of explaining the behavior of fluids. Fluid dynamics represents the trajectories of physical systems through a phase space defined over laws, initial conditions and boundary conditions. This theory thus commits us to a conception of aggregates of atoms as fluids and hence it commits us to an ontology of behaviors such as flowing absent at the level of the parts. Similarly, agential dynamics represents the hypothetically invariant trajectories of agents through ecological phase space defined over goals, repertoire, affordances. This theory commits us to a conception of organisms as agents and hence to an ontology of goals, affordances, repertoires and hypothetical invariance. In both cases, emergent properties play an indispensable explanatory role irrespective of the underlying conditions that realize this role.

In short, EA's silence about causal-location and hence the primitivist horn of the dilemma is a consequence of EA's commitment to a G1 criterion and the respective emergentist strategy of naturalization by behavioral ascent.

[3] In fact, since universality classes are multiply realizable (Batterman 2000), it is an advantage of EA that it makes no pronouncements about the physical constitution of agency. Agency may be realized in different ways, for example, in unicellular and multicellular organisms, and in aggregates of organisms such as colonies and swarms.

5.3.Causal and Explanatory Adequacy Attained

The scientifically emergentist criterion G1 not only gives a scientific model and justification for EA, but it also allows us to solve the dilemma by providing a strategy of naturalization that integrates the strategy of mechanistic descent represented by OA minus its essentialist definitional form, with the emergentist strategy of behavioral ascent represented by EA. The core idea is that naturalization is not a one-step procedure by causal strict definition, but a two-step procedure: First, explain the teleological-normative role that agency plays as a primitive concept in a scientific theory of organismal dynamics as per G1. And second, locate agency in the causal structure of the physical domain by non-circularly specifying the conditions that realize agency in the physical domain. The first stage specifies the nature and role of agency in irreducible phenomenological (ecological) terms. This specification takes the form of a primitive (circular) definition and has individuation force. In turn, the conditions that realize agency in the causal structure of the physical domain should be specified in non-agential strictly causal-mechanical terms. Crucially, this specification has no individuation force and hence does not take the form of a strict causal definition as per G2. This two-step emergentist strategy of naturalization seems to be conventional wisdom in condensed matter physics. As Goldenfeld and Woese (2011, 378-379) put it,

> "the modus operandi of condensed matter physics [is one in which] we regard a phenomenon as essentially understood when two conditions have been met. [...] identify the universality classes (i.e., categories) of interesting phenomena, and then to try and identify the likely realizations of them."

Applied to the case of biological agency, the idea is that EA phenomenologically identifies the organismal 'universality classes' by behavioral ascent in terms of the ecological concepts that constitute the agential dynamics of the system. In turn, OA identifies some of the 'likely realizations' by mechanical descent in terms of the causal concepts of closure and constraints that constitute the internal organization of the system. The robust and adaptive behavioral patterns that belong in the universality class of agential behavior are phenomenologically transparent and their dynamical profile is characterized by their hypothetical invariant modal profile. This strategy of behavioral ascent ensures explanatory adequacy while remaining agnostic about how these patterns are realized. In turn, since the underlying realizers are phenomenologically opaque, their identification requires transcending the phenomenology by mechanistic descent into the underlying causal regime characterized by closure of constraints, which locates agency in the thermodynamic order of the physical domain.

This two-step emergentist method of naturalization integrates EA and OA by allowing for complementarity through division of naturalistic labor. On the one hand, EA gives a dynamical phenomenological account of the teleological-normative nature and theoretical role of agency. This conforms to Jonas' (1966, 86) insight that "finalism is in the first place a dynamic character of a certain mode of existence, coincident with the freedom and identity of form in relation to matter." On the other, OA, minus its definitional requirement, gives a sub-agential causal-mechanical theory of how agency is realized by the causal organization of its physical constitution. This conforms to the original focus of the autopoietic conception of organizational closure as an account of "the *realization* of the living" (Maturana and Varela 1980, my emphasis). OA and EA are thus not two separate and thereby incomplete attempts at naturalizing agency, but two stages or aspects of a single and complete naturalization that can coherently satisfy both desiderata, therefore solving the dilemma.

6. Conclusion

As biology awakens from its gene-centric slumber and organisms re-emerge in all their irreducible complexity, the concept of agency is called to play a central theoretical role. But the naturalistic credentials of agency call for critical scrutiny—that is, for naturalization. I have proposed two desiderata: causal-location and explanatory indispensability, and compared two accounts: the Organizational or Constitutive Approach (OA) and the Ecological or Dynamical Approach (EA). I argued that while OA satisfies causal-location at the expense of explanatory adequacy, EA is explanatorily adequate but silent about causal-location—a dialectic that leads to a dilemma between causal reductionism and teleological primitivism.

My diagnosis was that each account is committed to different criteria of naturalism: causal reductionism is a consequence of OA's commitment to the stringent demands of G2 naturalism, scientific essentialism, and its strategy of naturalization by mechanistic descent. Teleological primitivism, by contrast, results from EA's commitment to the more lenient G1 naturalism, scientific emergentism, and its strategy of naturalization by behavioral ascent. Finally, I argued that G1 allows us to resolve the dilemma by integrating EA and OA, thereby satisfying both desiderata.

The result is a unified account according to which, biological agency is an ecologically specified universality class that plays a primitive teleological role in a phenomenological theory of organismal dynamics, while the underlying causal organization that realizes agential dynamics can be non-circularly specified in terms of closure of constraints, as part of a theory of how agency is physically constituted.

References

Arnellos, A., & Moreno, A. (2015). Multicellular agency: An organizational view. *Biology & Philosophy*, *30*, 333–357. https://doi.org/10.1007/s10539-015-9484-0

Barandiaran, X., Di Paolo, E., and Rohde, M. (2009) "Defining Agency: Individuality, normativity, asymmetry and spatio-temporality in action", *Adaptive Behavior* 17 (5): 367-386.

Batterman, R. (2000) "Multiple Realizability and Universality", *British Journal for the Philosophy of Science* 51: 115-145.

Batterman, R. (2005) "Critical phenomena and breaking drops: Infinite idealizations in physics", *Studies in History and Philosophy of Modern Physics* 36: 225-244.

Batterman, R. (2006) "Hydrodynamics versus Molecular Dynamics: Intertheory Relations in Condensed Matter Physics," Philosophy of Science vol. 73, 2006, pp. 888–904.

Batterman, R. (2017) "Autonomy of Theories: An Explanatory Problem," Nous, doi:10.1111/nous.12191

Batterman, R. W. (2021). *A middle way: A non-fundamental approach to many-body physics*. Oxford University Press.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies inHistory and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, 421–441. https://doi.org/10.1016/j.shpsc.2005.03.010

Bedau, M. (1992) "Where's the Good in Teleology?", *Philosophy and Phenomenological Research* 52: 781-805.

Bird, A. (2007). *Nature's Metaphysics: Laws and Properties*, Oxford: Oxford University Press.

Bray, D. (2012) "The Cell as a Thermostat: How Much Does it Know?", *Advances in Experimental Medicine and Biology* 736:193-199.

Burge, T. (2009) "Primitive Agency and Natural Norms", *Philosophy and Phenomenological Research* 79 (2):251-278.

Burge, T. (2010) *The Origins of Objectivity*, Oxford University Press.

Cartwright, N. (1983). *How the laws of physics lie.* Oxford University Press.

Christensen, W. and Hooker, C. (2000) "Autonomy and the emergence of intelligence: Organized interactive construction", in *Communication and Cognition: Artificial Intelligence*, 17: 133-157.

Cooper, J. (1987) "Hypothetical Necessity in Aristotle's Final Causes", in *Philosophical Issues in Aristotle's Biology*, ed. A. Gothefl and J. Lennox, 243-274. Cambridge: Cambridge University Press.

Corning, P. A., Kauffman, S. A., Noble, D., Shapiro, J. A., Vane-Wright, R. I., & Pross, A. (Eds.). (2023). *Evolution "on purpose": Teleonomy in living systems.* MIT Press.

Desmond, H., and Huneman, P. (2020) "The Ontology of Organismic Agency: A Kantian Approach". *Natural Born Monads: On the Metaphysics of Organisms and Human Individuals*, edited by Andrea Altobrando and Pierfancesco Biasetti, Berlin, Boston: De Gruter, pp. 33-64.

Di Paolo, E. (2005) "Autopoiesis, adaptativity, teleology, agency", *Phenomenology and the Cognitive Sciences* 4: 429-452.

Ellis, B. (2001) *Scientific Essentialism*, New York: Cambridge University Press.

Fulda, F. C. (2017) "Natural Agency: The Case of Bacterial Cognition." *Journal of the American Philosophical Association*, 1-22. doi:10.1017/apa.2017.5

Fulda, F. C. (2020) "Biopsychism: Life between Computation and Cognition." *InterdisciplinaryScience Reviews* 45:3, 315-330, doi: 10.1080/03080188.2020.1794381

Fulda, F. C. (2023). Agential autonomy and biological individuality. Evolution & Development. https://doi.org/10.1111/ ede.12450

Gibson, J.J. (1979) *The Ecological Approach to Visual Perception,* Boston: Houghton Mifflin.

Gilbert, Scott F. & Epel, David (2015). *Ecological Developmental Biology: The Environmental Regulation of Development, Health, and Evolution*. Sunderland, MA: Sinauer Associates. ISBN 9781605353449.

Gillett, C. (2016). *Reduction and emergence in science and philosophy*. Cambridge University Press.

Goldenfeld, N., and Kadanoff, L.P. (1999) "Simple Lessons from Complexity", *Science* 284: 87-89.

Goldenfeld, N., and Woese, C. (2007) "Biology's next revolution", *Nature* 25. DOI:1038/445369a

Gordon, D. M. (2023). Collective behavior in relation with changing environments: Dynamics, modularity, and agency. Evolution & Development. https://doi.org/10.1111/ede.12439

Hendry, R. F. (2008) "Two Conceptions of the Chemical Bond", *Philosophy of Science*, 75(5): 909–920.

Jaeger J. The Fourth Perspective: evolution and organismal agency (2023). In: Organization in Biology, Mossio M (ed.), Springer, Berlin.

Jonas, H. (1966) *The Phenomenon of Life: Toward a Philosophical Biology*, New York: Harper & Row.

Ke, B., Swarup, S., & Hwa, T. (2018). Metabolic fitness landscapes revealed by dual-omics analysis of adaptive evolution. Nature Ecology & Evolution, 2(5), 597–603.

Kim, J. (1989) "Mechanism, purpose and explanatory exclusion", in J. Tomberlin (ed), *Philosophical perspective 3: Philosophy of mind and action theory* (pp. 77-108), Atascadero, CA: Ridgeview.

Kim, J. (2005) *Physicalism, or Something Near Enough*, Princeton University Press.

Kirschner, M., Gerhart, J., and Mitchison T. (2000) "Molecular Vitalism", *Cell* 100 (1): 79-88.

Kirschner, M., and Gerhart, J. (2005) *The Plausibility of Life: Resolving Darwin's dilemma,* New Haven: Yale University Press.

Kripke, S. (1980) *Naming and Necessity,* Cambridge, MA: Harvard University Press.

Lewontin, R.C. (2000) *The Triple Helix: Gene, Organism, and Environment*, Harvard University Press.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*, 1–25. https://doi.org/10.1086/392759

McShea, D. (2012) "Upper-directed systems: a new approach to teleology in biology", *Biology and Philosophy* 27 (5): 663-684.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Springer.

Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge University Press.

Montévil, M., & Mossio, M. (2015). Biological organisation as closure of constraints. *Journal of Theoretical Biology*, *372*, 179–191. https://doi.org/10.1016/j.jtbi.2015.02.029

Moosavi, P. (2019) "From Biological Functions to Natural Goodness", *Philosophers' Imprint* 19 (51).

Moreno, A. & Mossio, M. (2015) *Biological Autonomy: a philosophical and theoretical enquiry* Springer.

Morrison, M. (2012) "Emergent Physics and Micro-Ontology", *Philosophy of Science* 79 (1): 141-166.

Morrison, M. (2014) "Complex Systems and Renormalization Group Explanations", Philosophy of Science 81 (5): 1144-1156.

Morrison, M. (2015) "Why More is Different", in Why More is Different: Philosophical Issues in Condensed Matter Physics and Complex System. Brigitte Falkenburg & Margaret Morrison (eds.). Berlin, Heidelberg: Springer.

Mossio, M., & Bich, L. (2017). What makes biological organisation teleological? *Synthese*, *194*, 1089–1114. https://doi.org/10.1007/s11229-014-0594-z

Nadolski, E. M., & Moczek, A. P. (2023). Promises and limits of an agency perspective in evolutionary developmental biology. Evolution & Development. https://doi.org/10.1111/ede.12432

Nagel, E. (1977) “Teleology Revisited”, *Journal of Philosophy* 84: 261-301.

Newman, Stuart. 2014. Review of ‘In Search of Mechanisms: Discoveries across Life Sciences’ by Carl F. Craver and Lindley Darden, University of Chicago Press, 2013, 228pp. In Notre Dame Philosophical Reviews, An electronic Journal. https://ndpr.nd.edu/news/in-search- of-mechanisms-discoveries-across-the-life-sciences-2/.

Newman, S. A. 2022. “Inherency and agency in the origin and evolution of biological functions.” *Biological Journal of the Linnean Society*, blac109. https://doi.org/10.1093/biolinnean/blac109

Nicholson, D. J. (2013) “Organisms ≠ Machines”, Studies in History and Philosophy of Biological and Biomedical Science 44 (4):669-678.

Noble, R., and D. Noble (2023) *Understanding Living systems*. Cambridge University Press.

O'Connor, T. (2021). Emergent properties. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2021 Edition). https://plato.stanford.edu/archives/fall2021/entries/properties-emergent/

Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). Niche construction: The neglected process in evolution. Princeton University Press.

Okasha, S. (2018) *Agents and Goals in Evolution*. Oxford University Press.

Papineau, David, "Naturalism", *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2023/entries/naturalism/>.

Pfennig, D. W. (editor). 2021. *Phenotypic Plasticity & Evolution: Causes, Consequences, Controversies*. CRC Press. Boca Raton, FL.

Psillos, S. (2002). *Causation and explanation.* Acumen.

Putnam, H. (1975) “The Meaning of ‘Meaning’”, *Minnesota Studies in the Philosophy of Science* 7: 131-193.

Rice, C. (2021) *Leveraging Distortions: Explanation, Idealization, and Universality in Science*, MIT Press.

Russell, E.S. (1945) *The directiveness of organic activities*. Cambridge University Press.

Salazar-Ciudad, I. Why call it developmental bias when it is just development?. *Biol Direct* 16, 3 (2021). https://doi.org/10.1186/s13062-020-00289-w

Sachs, C., and Nahas, A. (2023) "What's at stake in the debate over naturalizing teleology? An overlooked metatheoretical debate", *Synthese* 201(4): 1-22.

Sehon, S. (2005) *Teleological Realism: Mind, Agency, and Explanation*, MIT Press.

Shapiro, J. A. (2011) *Evolution: A View from the 21st Century*, FT Press.

Spinelli, J. B., Yoon, H., Ringel, A. E., Jeanfavre, S., Clish, C. B., & Haigis, M. C. (2021). Metabolic recycling of ammonia via glutamate dehydrogenase supports breast cancer biomass. Science, 371(6524), 80–85.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton University Press.

Simon, H.A. (1962) "The Architecture of Complexity", *Proceedings of the American Philosophical Society* 106 (6): 467-482.

Sommerhoff, G. (1950) *Analytic Biology*, Cambridge: Cambridge University Press.

Stich (1992) "What Is a Theory of Mental Representation?", *Mind* 101 (402): 243–61.

Strevens, M. (2005) "How Are the Sciences of Complex Systems Possible?", *Philosophy of Science* 72: 531-556.

Stovall, P. (2024) "The teleological modal profile and subjunctive background or organic generation and growth" *Synthese* 203 (3):1-37.

Strevens, M. (2008). *Depth: An account of scientific explanation.* Harvard University Press.

Sultan, S. E. (2015) *Organism and Environment: Ecological Development, Niche Construction, and Adaptation*. Oxford University Press.

Sultan, S. E., Moczek, A. P., & Walsh, D. (2022). Bridging the explanatory gaps: What can we learn from a biological agency perspective? BioEssays, 44, 2100185. https://doi.org/10.1002/bies.202100185

Thompson. E. (2007).*Mind in Life*, Harvard University Press.

Turner, S. (2017). *Purpose and Desire: What Makes Something Alive and Why Modern Darwinism Has Failed to Explain It*. HarperOne. San Francisco.

Tye, M. (1992) "Naturalism and the Mental", *Mind* 101 (403): 421-441.

Uller, T. 2022 Agency. Goal Orientation, and Evolutionary Explanation. In Evolution on Purpose (Open Access).

Varela, F. J. (1979). *Principles of Biological Autonomy*. North Holland.

Virenque, L., Mossio, M. What is Agency? A View from Autonomy Theory. *Biol Theory* 19, 11–15 (2024). https://doi.org/10.1007/s13752-023-00441-5

Walsh. D. M. (2008) "Teleology", In *The Oxford Handbook of Philosophy of Biology*, ed. Ruse, M.

Walsh, D.M. (2013) "Situated Adaptationism", in *The Environment: Philosophy, Science and Ethics*, Topica in Contemporary Philosophy Vol 9, Campbell, J.K.,M. O'Rourke and M. M. Slater (eds), MIT Press: 89-116.

Walsh, D. M. (2014) "Adaptation and the Affordance Landscape", In Pearce. T., Barker, G. and Desjardins. E. (eds). *Entangled Life: Organism and environment in the Biological and social Sciences*. Dordrecht: Springer pp. 213–236.

Walsh, D. M. (2015). Organisms, agency, and evolution. Cambridge University Press.

Walsh, D.M. (2018) Objectcy and Agency: Toward A Methodological Vitalism. In Nicholson, D. and J. Dupré (eds) *Everything Flows: Towards a process biology*. Oxford: Oxford University Press.

Walsh, D. M., & Rupik, G. (2023). The agential perspective: Countermapping the modern synthesis. Evolution & Development. https:// doi.org/10.1111/ede.12448

West-Eberhard. M.J. (2003) *Developmental Plasticity and Evolution*, Oxford University Press (New York).

Wilson, J. (2021). *Metaphysical emergence.* Oxford University Press.

Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*, Oxford Studies in the Philosophy of Science.