



# Normativity, Autonomy, and Agency: A Critical Review of Three Essays on Agency in Nature, and a Modest Proposal for the Road Ahead

Lenny Moss<sup>1</sup>

Received: 1 January 2024 / Accepted: 27 February 2024  
© The Author(s) 2024

## Abstract

Has the renewal of interest in the ostensible agency of living beings signaled an advance from a merely heuristic Kantian sense of purposiveness to an unequivocally, empirically grounded research program or are there as yet hidden tensions or contradictions in, for example, the organizational autonomy approach to natural agency? Can normativity be found to be immanent in nature but only beginning with the living cell or must a thoroughgoing naturalism find the seeds of normativity immanent throughout abiotic as well as biotic nature? Beginning with a brief exposition of Kant's influential treatment and recommendation for how to methodologically combine what he took to be the inevitable epistemological limit to explaining the origins of ostensible biotic purposefulness with the legitimate intentions of scientific research and explanation, this essay will critically engage with three recent essays that attempt to grapple with the preceding questions. Having putatively raised questions about the consistency and adequacy of each of the individual positions, the essay will attempt to move synthetically, drawing upon aspects of all three contributions, in the direction of a "cooperativity theoretic" approach to incipient natural normativity and agency.

**Keywords** Agency · Autonomy · Autopoiesis · Cooperativity · Far-from-equilibrium · Holobiont · Kant · Normativity · Organizational approach · Teleology

## Introduction

Agency in nature has become a hot topic across a number of disciplinary boundaries. Regrettably, perhaps, something like a reformulated renewal of the old "Two Cultures" divide (Snow 1959) still appears to hold sway. On the one side, now a spirit of *post*-humanism, with French philosophers Gilles Deleuze and Bruno Latour as leading lights, seeks to *explode* the human monopoly on agency with resulting putatively agentive shrapnel extending far and wide.<sup>1</sup> On the other side, driven largely by a regard for the apparent inherent "purposiveness" of the living organism, and inspired by Kant and the autopoietic theory of Maturana and Varela

(1980), theorists and philosophers of biology grapple with how to *expand* the franchise of human agency to include some or all living organisms within a naturalist framework. On the side of the former, all matter is deemed to have agentive status and questions of normativity have been largely sidelined if not eliminated. On the side of the latter, normativity is taken to be an irreducible aspect of organismic agency even if how to reconcile this with naturalism is a work in progress. For all intents and purposes, dialogue between these two frameworks has been nonexistent.

Living beings have perennially been seen as existing as agents acting on behalf of their own continued existence and well-being. For Aristotle the living being is thus its own final cause (or *telos*) and in its natural self-purposiveness constitutes the most fundamental "substance" or "primary ousia" in a nature that ontologically is invested with immanent purposiveness (Moss 2017). For Kant, writing under very different ontological/metaphysical assumptions, recognition of the self-purposiveness of "organized beings" was also an unavoidable, albeit *subjective*, judgment of human reason (Kant[1790]2000). Kant offered natural science a "methodology of teleology" that would enable "objective science"

<sup>1</sup> Readers may be interested in my discussion of this trend in Moss (2017).

✉ Lenny Moss  
dcimorse@icloud.com

<sup>1</sup> Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México, Mexico City, Mexico

to go as far as possible in bringing living processes under its explanatory ambit just so long as it was understood that the purposive organization of life had to be taken as a background given that could not itself be scientifically accounted for. The question of agency in nature has again become increasingly prominent, and coming from various antipodal quarters of the academy, but there is much left to be desired so far as critical encounters between very different views. *Biological Theory* has taken some initiative in providing a forum for advancing theoretical and critical expositions and debates on the topic of agency in biology. Desiderata for any such endeavor could and should include the following: (1) expanding the scope of views that are engaged in active conversation, (2) promoting critical engagement of said views, (3) steering away from temptations toward settling into insular practices of “normal science,” (4) promoting the union of theoretical and empirical investigation with philosophical reflection and clarity, and (5) promoting historical as well as synchronic reflection on our current state of understanding. The following, for example, may be indicative of the disparate positions that various investigators may want to defend and bring into debate: (a) we have essentially made no further progress since Kant in fully naturalizing our understanding of agency and never will absent a new revolution in our basic understanding of nature; (b) we are still within a Kantian framework but our advances in science allow us to extend the boundary between what we can explain scientifically and what we must still bracket as a purposive given; (c) Darwinism has long since done away with any problem in accounting for natural agency; (d) advances in the physics and chemistry of far-from-equilibrium excitable soft condensed matter has fully accounted for natural agency (or shows promise for being able to do so); (e) agency simply is immanent throughout nature, it is only the legacy of humanism that has impeded our view, and organisms (including ourselves) are thus only transient constellations of an agentiveness of matter that is ubiquitous and always already there. In the following, I will be reviewing and commenting on essays by Okasha (2023), Pickering (2023), and Virenque and Mossio (2023).

### **In the Background: Kant's Critique of the Power of Judgment**

Inasmuch as Kant's formulations in his *Critique of the Power of Judgment* (Kant [1790]2000) have either directly or indirectly set the stage for what's become the most prevalent approach (or approaches) to the concept of agency in biology, it is worth adumbrating the main thrust of Kant's argument and legacy. I will do this through an annotated presentation of selected key passages from his text, beginning with the headings from the sections where each passage

is located. It is worth noting that in Kant's usage “critique” refers to the specification of the conditions of possible understanding while “dialectic” refers to trajectories where, in Kant's view, reason leads to contradictions and illusions.

#### ***Critique of the Teleological Power of Judgment.***

##### **§ 66. On the principle for the judging of the internal purposiveness in organized beings.**

This principle, or definition, states: An organized product of nature is that which everything is an end and reciprocally a means as well. Nothing in it is in vain, purposeless, or to be ascribed to blind mechanism of nature.... One can call this principle a maxim for the judging of the inner purposiveness of organized beings.

Kant goes on to explain that anatomists seeking to understand structure (and we can project ahead to physiologists as well) have used this principle as a maxim, or a rule of thumb, in conducting their research. The thrust of the argument here is that the judgment of purposiveness is *unavoidable* in the study of the living organism. It is nothing less than a condition of possibility for the scientific study of life. Kant famously elaborates on this with respect to the apparent circularity of causes—parts acting for the sake of the whole, the whole acting for the sake of the parts—a state of affairs that we can only cognize by perceiving the “organizing being” as existing as an end or purpose unto itself.

#### **Dialectic of the Teleological Power of Judgment.**

##### **§ 69 What is an antinomy of the power of judgment?**

The *determining* power of judgment... merely subsumes under given laws or concepts as principles.... But the *reflecting* power of judgment is supposed to subsume under a law that is not yet given and is in fact only a principle for reflection on objects for which we are objectively entirely lacking a law or concept of the object that would be adequate for the cases that come before us.

Kant distinguishes between two forms of judgment, determining judgment and reflective judgment. In the case of a determining judgment the nature or identity of some phenomenon has been determined by an established law. Something falls off our desk and we take it to be an object with mass that has been subjected to the force of gravity. If there is a breeze in our office and something floats away from our desk and flutters in the air before falling we may judge it to have a low mass and be subject to the force and buoyancy of the breeze. But if something appears to fly off our desk on its own and perhaps head toward a window, we do not have a determining judgment but rather must resort to a judgment

of reflection that, failing to invoke an already established principle, reflectively judges that the object must be acting for its own purpose. Perhaps we can borrow a bit of everyday phenomenology to help makes sense of Kant's subsequent point. Kant is about to introduce the idea that we are faced with an ostensible, if ultimately only an ostensible, antinomy of conflicting maxims. Returning to our desk, would it not be phenomenologically apt to say that when something falls off our desk our first inclination is to be guided by the maxim of our understanding of physical causation and that it is only when the adequacy of this maxim appears to fail, however subconsciously this may occur, that we resort to the reflective judgment that a purpose outside of our own has come into play? The ostensible antinomy of judgment is then expressed by Kant as a thesis and an antithesis.

**"Thesis** All generation of things is possible in accordance with merely mechanical laws.

**"Antithesis** Some generation of things is not possible in accordance with merely mechanical laws."

Kant proceeds to proffer that the claim or appearance of an antinomy is based upon a confusion of what amount to two different functions of reason and he thereby gestures toward the sense that these alternative maxims may play complementary roles in practice. He proceeds to spell out the nature of this practice in a lengthy appendix as follows.

*Methodology of the Teleological Power of Judgment.*

**§ 80. On the necessary subordination of the principle of mechanism to the teleological principle in the explanation of a thing as a natural end.**

It is thus rational, indeed meritorious, to pursue the mechanism of nature, for the sake of an explanation of the products of nature, as far as can plausibly be done, and indeed not give up this effort because it is impossible *in itself* to find the purposiveness of nature by this route, but only because it is impossible *for us* as humans—since for that an intuition other than sensible intuition and a determinate cognition of the intelligible substratum of nature, which could furnish the ground for the mechanism of the appearances in accordance with particular laws, would be necessary, and this is entirely beyond our capacity.

If, therefore, the investigator of nature is not to work entirely in vain, he must, in the judging of things whose concept as natural ends is indubitably established (organized beings), always base them on some original organization, which uses that mechanism itself in order to produce other organized forms or to develop its own into new configurations (which,

however, always results from that end and in conformity with it).

Kant exposes here his basic approach to overcoming the apparent antinomy between the determinative (mechanistic) maxim and the reflective (teleological) maxim of judgment. He proposes that so long as we take the original purposive organization of the life form as a given that can't be derived from mechanistic first principles, we are justified, indeed behooved, to take mechanistic analysis as far as we can in explaining how mechanisms are used to realize the ends of life. We should note that Kant has not ruled out the possibility of "new configurations" being produced so long as they are in tune with the purposiveness of the organism (more on this below). Kant also conditions this "methodology of teleology" on what he takes to be the limits of human cognition and not the limits of any possible cognition. Does this mean that a self-learning artificial system may be able to acquire the capacity for "determinative judgments" that overcome the mechanistic versus teleological distinction? Or perhaps our sciences have already so sufficiently expanded our stock of basic objective laws and principles such that we can at least revise the boundary between determinative and reflective judgments (as represented in the enumerations offered above) if not entirely abolishing the distinction?

**§ 81. On the association of mechanism with the teleological principle in the explanation of a natural end as a product of nature.**

No one has done more for the proof of this theory of epigenesis as well as the establishment of the proper principles of its application... than Privy Councilor Blumenbach. He begins all physical explanation of these formulations with organized matter. For he rightly declares it to be contrary to reason that raw matter should originally have formed itself in accordance with mechanical laws, that life should have arisen from the nature of the lifeless, and that matter should be able to assemble itself into the form of a self-preserving purposiveness by itself; at the same time however, he leaves natural mechanism an indeterminate but at the same time also unmistakable role under this inscrutable principle of an original organization, on account of which he calls the faculty in the matter in an organized body.. a *Bildungstrieb* [formative drive].

The 18th century began with what we would now call "preformationism" as the predominant approach to understanding embryonic development. New theories of epigenesis arose largely inspired by the fact of Newton's theory of gravity establishing as scientifically legitimate what

had been previously deemed an “occult force” acting at a distance (and as such contrary to orthodox mechanism). Debates began to rage between “evolutionists (preformationists)” and new advocates, such as Friedrich Caspar Wolff, of a “Newtonian vital force” in organismic development. Already in earlier “pre-critical” writings, Kant had devised a mediating concept in his defense of monogeny (the doctrine of the descent of all humans from a single original pair) versus polygeny (the doctrine that different human races were derived from different original ancestral pairs). In championing the former, Kant assumed responsibility for being able to account for the diversity of human phenotypes deriving from an original pair. His solution was to posit an original stock (*Stamm*) of germs or seeds (*Keime*—a concept used by post-*emboitement* preformationists) and proclivities or facilities (*Anlagen*—a concept first introduced into biology by Kant in order to capture the wider spectrum of possible phenotypic outcomes) resulting in what he came to refer to as “generic preformationism.” The claim was that humanity began with an original stock of developmental/formative potential that could become specialized in response to specific environmental pressures, i.e., darker complexion in more equatorial regions, flatter noses and glare-resistant epicanthal eye folds in the face of frigid temperatures and high glare reflecting off snow, and so on. But unlike the Darwinian approach to adaptation of nearly 100 years later, in Kant’s model, residual developmental potential diminished with each specialization. Kant went on to further articulate this model with empirically based generalization about the results of hybridization between specialized forms that resulted in intermediary characteristics.

Johann Blumenbach by the 1780’s had become a leading figure in areas of comparative biology and physiology and the acknowledged founder of physical anthropology. Kant, as above, looked to Blumenbach as a sympathetic interlocuter in advancing a methodology of teleology. Inasmuch as Blumenbach’s approach to epigenesis of form and adaptation did not assume the appearance of form *ex nihilo* but always presupposed “an original organization” possessed of “a self-preserving purposiveness,” Kant could see in the *Bildungstrieb* a guidepost for formulating a methodology of teleology. Kant embraced in Blumenbach’s concept of the *Bildungstrieb* what he took to be a way of mediating between mechanistic and teleological principles without collapsing the distinction between them. The *Bildungstrieb* was a way of objectifying (albeit heuristically for Kant) the capacity of the organized cell/organism to direct its internal mechanisms for the sake of its well-being, i.e., for a purpose. In promoting and advancing contemporary discussion about agency in life, we would like to see the same degree of philosophical rigor and precision that we find in Kant in contemporary discourse. We will want to ask whether new

concepts that are being deployed for approaching questions of biological agency are, explicitly or not, being used to mediate between teleological and mechanistic approaches? Are such concepts being used in reflective and heuristic mood or meant to constitute a determinative principle, and if the latter is such usage warranted? Might it be the case that some successor to the *Bildungstrieb* may, for example, constitute a kind of mediating placeholder, that aspires to hold open a conceptual space for a transitioning from a heuristic concept to a new scientific principle (and thereby obviate the need for the mechanism/teleology binary)? The idea of “autonomy” in particular has gained widespread usage in the biological agency literature and prominence in ongoing research programs. We will want to be mindful of where autonomy and other recent concepts stand in relation to this background context and explicate what possible implicatures they bring with them.

### Three Essays on Agency in Nature

In his contribution to the discussion, Samir Okasha’s paper “The Concept of Agent in Biology: Motivations and Meanings” (Okasha 2023) attempts to provide what his title suggests. His strategy is to first identify what the motivations are for approaching the question of agency, then enumerate the concepts of agency that are discernable in contemporary usage, and finally try to find the best fit between a meaning of agency and a motivation for approaching the topic. Ostensibly, Okasha would appear to be engaged in a kind of analytic clarification that does not mean to evaluate whether progress has been made in accounting for agency in nature but rather only to clarify what concept of agency would be relevant to such an evaluation given an underlying motivation. Whether in fact one can isolate a motivation from an assumed meaning would be one possible underlying factor complicating the achievement of his intentions.

Okasha distinguishes between two motivations. In the first case he draws upon the idea that agency pertains to organisms in a way that it doesn’t pertain to either higher or lower levels of biological formations. In so doing, he already calls attention to the legacy of Francesco Varela (1979, 1981) that designates autonomy as the concept that describes what is distinctive of organisms. Okasha goes on to append as an additional motivation that which seeks to oppose genetic reductionism and restore a sense of integrity (or one might again say autonomy) to the organism. If, however, our interest is in seeking to understand the sense of agency in nature more broadly might it preempt an adequate inquiry by already presupposing that agency is relegated to the organism? If we are already specifying agency as that which is specific to the organism isn’t there then a circularity in seeking to pick out what we mean by agency? Might

it not be the case, for example, that organismic agency is a particular form of natural agency and that without a grasp of the more general sense of agency we will stagnate in our effort to better understand the particular? Okasha refers to this first motivation as an Organism-as-Agent Theory or OAT. His second motivation by contrast does not seek to take immanent agency in nature in a realistic fashion but rather to see the language of agency as a useful heuristic for approaching the consequence of evolution by natural selection. In analogy with Dennett's intentional stance (Dennett 1989), this motivation for agency talk just seeks the benefits of utilizing familiar intentionalistic speak in an "as-if" mood when "we all know" that it's really natural selection that is doing the heavy lifting. Okasha refers to this latter motivation as the Organism-as-Agent Heuristic or OAH.

Moving on to the topic of different concepts of "agency" Okasha suggests four. First is a minimal sense, derived from Fred Dretske, that simply distinguishes between acting or being acting upon. This sense doesn't distinguish between whole organisms and other biological formations such as organs or organelles nor between artifacts and living things, nor is it clear whether natural inorganic phenomena such as rivers (or icebergs) carving valleys would also count. Second, and on the opposite end of the spectrum, is the intentional concept of agency, that presupposes and requires the fact of intentional mental states that prefigure an action. Okasha understands this concept of agency to only be apropos of humans. Third is a concept Okasha attributes to AI, thus dubbed the "intelligent agent," in which an agent is characterized by specific and flexible responses to an environment, albeit that need not entail mental intentions. A thermostat would count as an agent under this concept. Lastly is the concept of a utility-maximizing agent that Okasha refers to as a "rational" agent following the lead of an economist's rational choice model. In this case the concept refers specifically to the agency of always maximizing utility (again without presuming intentional mental states).

In mapping concepts onto motivations (or theories) it would be easiest to begin with OAH. Okasha flirts with the idea that the intentional concept is a good fit inasmuch as it uses as-if language for nonhuman organisms to replace psychological states in describing adaptive behavior on the part of organisms (i.e., behavior that was selected for). He then refers to a "Grafen-style" version of OAH which would appear to be closer to the economic agent who is maximizing utility-cum-fitness. Given that the whole gist of the OAH is to treat the appearance of agency in organisms as only the result of natural selection for higher fitness I don't know that I see any substantive difference between these two as the "as-if" intentions of the intentional concept are relegated to only that which can be rationalized from a Darwinian (or neo-Darwinian) point of view. In other words,

if you treat the intentional concept of agency only in an as-if vein, and from a Darwinian perspective, does it not come out as being a lot like, or simply identical with, a non-intentional utility-maximizing concept (whereby utility just means fitness maximization)?

Coming back to his OAT Okasha appears to find himself taking a "least of all evils" perspective. If the intentional agent concept is too demanding, the minimal agent concept is not demanding enough and the rational agent concept is not an entirely apposite fit, then the intelligent agent concept would seem to be the best choice. Be that as it may, the intelligent agent concept as defined fails to meet the desideratum of distinguishing between organisms and other entities, biological or artificial, that may be empowered to engage in flexible and goal-oriented responses to environmental fluctuations and contingencies. Okasha goes on to suggest that the underlying reason for the lack of a fully adequate mapping is that "none of the four concepts captures the idea of autonomy, which intuitively is what distinguishes organisms from other flexible goal-directed systems in nature" (Okasha 2023). This brings us back to our earlier concern over whether by initially associating the realist understanding of agency with that which is uniquely distinctive of organisms one was begging the question of the agency concept and preempting a wider inquiry into the nature of natural agency. In the last section of this essay, where we revisit the question about progress since Kant, we will look at the benefits and liabilities of delimiting the scope of agency to within a framework of autonomy.

In his essay "What is Agency? A View from Science Studies and Cybernetics" Andrew Pickering provides a lucid window onto the perspectives of those on the "other side" of the latest "Two Cultures" divide (Pickering 2023). Rather than delimit the province of agency to organisms from the get-go (which some may see as a projection from a tired humanism), Pickering begins with the assumption that nature as a whole is a "dance of agency." The empirical warrants for his claims stem from science-studies analyses of physics laboratory research as well as more recent work on the history of British cybernetics. In the case of the former, and indeed germane to the experience of experimentalists more generally, unexpected outcomes are far more the norm than the exception up until a system is secured in which "the flux of becoming" is stabilized and made predicable, and only then can the agency of matter be backgrounded and human versus object relations established. Pickering refers to these as the finding of "islands of stability" (others such as philosopher Hubert Dreyfus have referred to it as "beating nature into shape").<sup>2</sup>

<sup>2</sup> Personal communication, 1983.

As we saw in Okasha's paper above, inquiry into agency began with the assumption of biological agents and this is indeed the norm for studies on the theory and philosophy of biology side of the divide. Pickering, by contrast, wants to challenge this assumption in favor of the idea that agency is universally distributed and pervasive. He suggests that we thereby "think of agents, as contingent fractal islands of stability in the flux of becoming" (Pickering 2023). By fractal he means "built up in a regress of sub-agents."<sup>3</sup>

In his essay, and possibly for the first time, Pickering attempts to bring his account of "minimal agency" (a term borrowed from Okasha) to bear on issues of agency in biology. His first movement, recollecting the history of cybernetics, is to see some form of reentry or recursion as characterizing an assemblage of parts that transition from purposeless to purposive agency. The thermostat is his canonical example. "The overall plan of goal-orientation is, then, an appropriate organization of parts coupled to an environment so as to create a closed loop in the dance of agency—here of a multiplicity of minimal nonhuman agents acting together, and jointly constituting a purposive higher-level agent (the thermostat)" (Pickering 2023). With talk of organization and closed loops, Kant, autopoiesis (acknowledged by Pickering), and autonomy theory spring to mind. But is the idea of the emergence from lively matter to an "'organic' island of stability" sufficient to account for how nonnormative agency becomes normatively structured agency? Does offering an alternative ontology for understanding physical phenomena allow us now to obviate any need for mediating between mechanistic and teleological principles? In approaching questions of normativity (if not in so many words) Pickering suggests that self-learning systems, whose pathways of learning have become inscrutable (as we would now say about chatbots that can produce different outputs to the same question ad infinitum) are examples of nonorganic systems with emergent normativity. That said, one may yet want to protest that a transition from nonnormative to normative systems still requires (normative) human intervention, and is thereby not spontaneous, however much an artificial system may *thereafter* be able to spontaneously bootstrap itself into a considerably more complex or unpredictable normative system.

The essay by Virenque and Mossio "What is Agency? A View from Autonomy Theory" provides a window onto what has become a widespread research community focused on the idea of "autonomy," "organizational closure," and the "organizational approach" to understanding agency (Virenque and Mossio 2023). Virenque and Mossio (V&M)

are explicit in understanding the autonomy-centered approach to agency, to which they contribute, as originating and following the lead of Kant: "the theory of autonomy submits that living beings possess a distinctive organization that, to use the famous Kantian formula ([1790]1987), can be legitimately said to be 'cause and effect of itself'" (Virenque and Mossio 2023). Echoing Kant's account in his section "the Critique of the Teleological Power of Judgment," V&M reiterate the claim that the continuous sustaining and producing of the organization of the organism, that which enables the parts to function on behalf of the whole, is the inherent and natural purpose of the organism. What V&M do not take from Kant is the epistemological distinction between determinative and reflective judgments. While using language that Kant referred to as teleological, V&M (and we have no reason not to assume their many collaborators in the autonomy research community) assert that their formulations are fully naturalized and scientifically legitimate. Indications of some misreading of Kant notwithstanding, whether the collapse of the determinative versus reflective distinction has become an enabling pathway for better grasping biological agency, or has led to serious problems, needs to be evaluated on its own terms. While returning to the larger question of whether progress has been made in grasping natural agency awaits our concluding section, at present we will focus on the further logic of V&M's formulations.

We suggested earlier that Kant adopted notions such as the *Bildungstrieb* (formative or developmental force) as a kind of intervening concept that mediates between determinative judgments of constitutive mechanisms and regulative concepts of purposefulness. It would follow, if not made explicit, that contemporary investigators who wish to eschew any form of epistemological dualism would attempt to advance the equivalent of intervening concepts, albeit such as are entirely within the ambit of established scientific principle. V&M (2023), and many collaborators, have taken a direct implication of the "mutual dependence among the functional parts of a purposive organization" to be that of "organizational closure." The warrant for this inference that V&M offer is that the parts of the system are subject to degradation and therefore that there must be organizational closure to allow for the replacement of parts without resulting in ongoing changes/mutations to the organization as degraded parts are replaced. They then proceed to suggest that organizational closure "implies thermodynamic nonequilibrium," as a constant input of energy is required to replace parts while strictly conserving organization. Procuring energy is then established as the basic structure of the relationship between an autonomous system and its environment, which then issues into a definition of agency.

<sup>3</sup> As exotic as this may sound, studies in the neurosciences of human cognition have moved in the direction of understanding the experience of individual identity agency as the product of any number of cognitive subagents (such as quasi-independent right and left hemispheres).

Agency, in other words, consists in the (inherent) interactive dimension of organizational closure, in those functional capacities of a living being devoted to purposively governing the relationship with the environment...The theory of autonomy offers, therefore, a perspective from which agency can be understood as a behavior performed for a reason, directed toward an intrinsic goal, which is the continued existence of the system's self-determining organization, through an incessant interaction with its external environment. (Virenque and Mossio 2023)

Through a stepwise series of implications, V&M have come to a definition of agency, albeit by way of certain constraints or commitments that they then see as raising some possible difficulties. We will also want to consider to what extent these difficulties may or may not be consequences of effectively eschewing the teleological heuristic and conferring an exclusively "determinative" status upon concepts of organization, organizational closure, autonomy, and so on.

Virenque and Mossio are not satisfied to end with the exposition of agency and autonomy they have put forward because there are challenges from within their own research community that warrant consideration. They begin by recalling Di Paulo's (2005) criticism that an autopoietic system cannot account for adaptivity. They then go on to affirm that all living systems are in fact adaptive systems (and thus subjects of adaptive agency) which entails "the possibility to shift to different and new organizational regimes." While one would think that this would appear to problematize the account of autonomous agency that they have proffered, they go on to circumvent the problem, by assertoric fiat it would seem:

As Moreno and Mossio (2015, p. 104) point out, "Auto-nomy here is not just the maintenance of the current condition of existence, but the fact of promoting its own existence on behalf of a more fundamental (and less contingent) identity." The identity of the system is less contingent because adaptive agency enables (continuously) changing its own current organization and behavior to keep existing. (Virenque and Mossio 2023)

If we recall the earlier stage of their account, sustaining the organization of the cell qua organism was put forward as the ultimate end to which all the parts of the cell were functionally subordinated to. In the absence of any determinative versus reflective distinction, conferring self-purposive constituting prowess to the organization of the cell is being treated as an empirically accountable scientifically realist proposition. V&M and others have endeavored to

further harden the characterization of organization from that initially put forward in autopoiesis by advancing the characterization of organizational closure as an empirically accountable phenomenon. In their adaptation-problem inspired recourse to the quote from Moreno and Mossio above, we see that the word autonomy has been hyphenated in order to make conspicuous the etymological sense of autonomy as meaning self-norming. Suddenly, it seems the problem of addressing the adaptation problem is solved by etymological fiat because autonomy simply implies the ability to create its own norms and if the norms of the cell are invested in its organization then autonomy implies being able to change its organization. Put forward as a determinative judgment, what scientific principle would account for the ability of an organization to normatively transform itself into another norm enforcing organization? Is this really an empirically accountable proposition or a reflective teleological judgment? If one treats "organization" in a realist/mechanist fashion then asserting an ability to change its organization as "adaptive agency requires" would seem to suggest an internal contradiction. What is the mechanism by which an organization that maintains itself is also able to transform itself? What is the warrant for claiming that such an "organization" is understood in a naturalized fashion?

V&M end their paper by calling our attention to one more putative conundrum. Inasmuch as an adequate characterization of any life-form would be such as to be a subject of adaptive agency, and adaptive agency implies a capacity for making sense of one's environment (without which "adaptation" would be meaningless), we must see the roots of cognition in elementary environmental sense-making. But inasmuch as we take cognition to transcend the limits of pure self-interest and rather aspire toward interest-neutral objectivity, just how and when does the trajectory of entities theoretically construed as self-sustaining systems of "looking out for number one" (my words) enter into the exercise of objective cognition?

## Discussion

Having now introduced the views of the three essays under consideration, we can presently reorient "the narrative" to the extent of relating the papers to each other and to more general and historically contextualizing questions that we began to adumbrate in the introduction. Have we, for example, gone beyond the Kantian "methodology of teleology" and if so, are the results warranted and commendable?

Pickering's view of the status of natural agency, as he has confirmed, draws upon Okasha's "minimal" sense or meaning of agency and also affirms my prerogative proposition "e," i.e., "Agency simply is immanent throughout nature, it

is only the legacy of humanism that has impeded our view, and organisms (including ourselves) are thus only transient constellations of an agentiveness of matter that is ubiquitous and always already there.” By the terms of the minimal view, however, and Pickering’s appropriation of it, agentive materiality is nonnormative (it just acts) and yet somehow becomes normatively structured within so-called islands of stability (organic or otherwise). This raises the challenge as to how normativity can arise from the absence of normativity. But must we assume that natural materiality lacks normativity? This is certainly a fundamental, and perhaps THE fundamental, commitment of that metaphysical sea change we refer to as “The New Science” of the Renaissance. So what would a plausible alternative look like? Pickering’s point of departure insofar as according agency to natural materiality in general is that of the standpoint of the laboratory experiment, with the detection of free quarks being his prime exemplar. Agency, for the experimenter, then shows up as a kind of resistance, as anomaly, and as at least an initial obstacle to being able to redeem one’s hypothesis and declare a principle, i.e., a determinative norm of nature. Ironic as it may be, the experience of the nonnormativity of agentive nature is thus an artifact of the investigator’s failure to, at least initially, “beat nature into shape.” Residual anthropocentrism is thus key to the apprehension of nature as nonnormatively agentive. But what then would the alternative look like?

To avoid the conundrum of accounting for something coming from nothing, we only need to identify a something, no matter how rudimentary, that can be seen as a source from which something more can arise. I would suggest that the most rudimentary expression of normativity, or perhaps protonormativity, in nature would be that of incipient cooperativity. And it is easy enough to imagine that unless one is specifically searching for signs of incipient cooperativity it is most likely to show up as resistance, as untamed nature, as nonnormative agency. To be able to recognize incipient cooperativity we need to be open to it, i.e., to bracket our inclination to beat nature into a shape that conforms to our prior determinative judgments. The dynamic behavior of liquid water (the wellspring of all life as we know it) would be a particularly apposite venue for approaching natural materiality from a cooperativity-theoretic point of view. It is, for example, well established that organized cooperative structures (even quantum alignment?) spontaneously form at interfaces between water and air or other nonaqueous media (see Pollack 2013 and Ho 2012). I would suggest that this is only the “tip of the iceberg” when it comes to allowing ourselves to see autochthonous cooperativity, no matter how ephemeral, in aqueous (and other) media. Where the expression of spontaneous cooperativity has been well attended to and thematized has been in the cases

of reproducible self-organization in far-from-equilibrium phenomena such as the Belousov-Zhabotinsky reaction, made well known especially by Ilya Prigogine’s model of the “Brusselator” (Prigogine and Stengers 1984; see also Winfree 1984). Just as the far-from-equilibrium state can be viewed as an amplification of the always somewhat removed from equilibrium state of matter as we know it, including soft or fluid condensed matter, could it be likewise the case that frank self-organizing systems are amplifications of incipient cooperativities that are ubiquitous if only we (experimentally) open our eyes to seeing it? We will shortly want to consider how and why incipient normativity qua cooperativity has come to be construed, in much of the biological agency literature, to be inimical to autonomy.

Samir Okasha, whose essay was oriented toward a descriptive clarification as opposed to a critical assessment of current research on that status of agency in biology, offered us two approaches that he dubbed the OAH and OAT. Inasmuch as the use of teleological language within OAH (organism as agent heuristic) would be for the purpose of facilitating our insight into the adaptive achievement of natural selection, we can easily conclude that OAH is consonant with our proposition “c”—“Darwinism has long since done away with any problem in accounting for natural agency,” i.e., an ostensibly eliminative approach to natural agency as a problem. When it comes to OAT (organism as agent theory) we found that none of the four familiar senses of agency met the desiderata that Okasha set out and that OAT had to be cashed out in terms of the idea of autonomy as that which is the exclusive attribute of organisms. This then leads us back to the essay by Virenque and Mossio that is all about putting forward autonomy theory as the enabling pathway for understanding agency in nature.

We have already raised concerns about whether Okasha’s assumption that the motivation for investigating the nature of biological agency from a realist motive presupposes the unique status of the organism (perhaps primarily thought of as the living cell) in that regard (while disregarding the question of agency at other levels). Okasha was not defending that view but only attempting to characterize what appears to be the predominant view at present, much of which has been directed toward the “autonomy” concept. Virenque and Mossio, expositors of the autonomy view, have already acknowledged possible problems of insularity beginning with criticisms of the antecedent autopoiesis concept pertaining to adaptability. These concerns have become significantly amplified in a recently funded grant proposal by Leonardo Bich and coworkers.<sup>4</sup> Bich has been one of

<sup>4</sup> The project, entitled “Autonomy—Fleshing Out Autonomy Beyond the Individual” ([https://www.ias-research.net/wp-content/uploads/2020/09/Autonomy\\_project-description.pdf](https://www.ias-research.net/wp-content/uploads/2020/09/Autonomy_project-description.pdf)), was submitted in 2019 to the Spanish Ministry of Science and Innovation and



the most prolific contributors to the biological autonomy research program but in that recent IAS-Research (University of the Basque Country) grant proposal expressed the need to meet challenges to the insularity of an individualistic autonomy perspective that have been posed by many relatively recent developments in the sciences from holobiont theory, extended cognition, collective forms of autonomous agency, “to the pressing need to develop tools to understand sustainability and self-governance in biological and social organization.” Bich and colleagues have proposed to meet these challenges through moving from the standpoint of autonomy to that of an “outonomy” although preliminary results of this project have yet to surface. Both the scope of research engagement with the autonomy perspective and indications of cracks in its armor both from within and without the autonomy research community speak to the warrant for a thorough (re)assessment going back to its roots in, and departure from, the Kantian background.

We had earlier made a passing reference to an indication in the Virenque and Mossio paper of a misreading of Kant. Early in their paper, in a section specifically aimed at “Naturalizing Agency from the Perspective of Autonomy,” V&M tell us that,

Living beings are autopoietic because the concerted activity of their parts results in their reciprocal continued production over time: consequently, the whole system is cause and effect of itself. *Pace* Kant, however, no force is at play here: the *organization* of the parts is such that they collectively contribute to their own existence. (Virenque and Mossio 2023)

V&M are here referring to Kant’s embrace of Blumenbach’s notion of the *Bildungstrieb*. Kant was well aware that we do not have scientific principles for explaining how “an organization” can result in a system becoming cause and effect of itself. Kant was very clear that “if the investigator is not to work entirely in vain” they must begin with a reflective assumption that somehow, by means unknown, we have an organization that confers a purposiveness onto otherwise garden-variety parts that left to their own devices would simply conform to the purposeless laws of physics in their activity. Kant’s embrace of the term *Bildungstrieb* (or formative force) to describe the conferral of purposiveness onto the parts was entirely, and uncontroversially, intended in an “as if” vein. One of the advantages of using an “as if” concept of a force (as if it were able to confer purposiveness onto the parts and thereby the whole) as a placeholder for dynamics that we do yet understand in a “determinative

way” is that it defers any need for imposing restrictions that foreclose other possibilities. And isn’t the absence of this benefit exactly what has led to the aporias of the self-enclosing “autonomy” paradigm?

In attempting to “harden” a heuristic placeholding concept into a putative empirically accountable law or principle one is faced with a dilemma. Opting for a particular mechanism as the basis for explaining a phenomenon brings with it both positive and negative implications. There are phenomena that both follow from said mechanism and those that are specifically prohibited by it as well. While nobody would or should pose objections, in principle, to the intentions of “hardening” heuristic concepts in order to achieve a fully serviceable and empirically accountable explanatory system, one must also be cognizant and responsible for the trade-offs and liabilities that are consequent upon moving in a particular direction. V&M (2023) tells us that “agency is conceived within the theory of autonomy in a fully naturalized way, as soon as the underlying causal regime—organizational closure—is deemed to meet the epistemological standard of natural science (a point we take for granted here).” Let us then for a moment reflect upon the rendition of the concept of “autonomy” as it has been shaped by the “underlying causal regime” of organizational closure.

Autonomy (as highlighted by V&M as “auto-nomy” in quoting Moreno and Mossio 2015) etymologically means self-norming. Kant very specifically does not bring the concept of autonomy to bear in relation to organized beings as he sets his concept of autonomy in relation to a very high bar. Autonomy, or normative self-legislating, for Kant is restricted to “rational” beings whose reason demands recognition that for any maxim of action to be normatively justified one must recognize that it must likewise be what one would will for any other rational being in the same situation. It thus follows for Kant that inasmuch as one does not wish to be treated as merely another’s means to their own end, that to be autonomous likewise means that one will not treat another rational being as a means to an end. There is then some irony in encountering a rendition of autonomy, with putative roots in Kant, that has found its way to identifying autonomy with all and only perseveration of its own organization, i.e., with “looking out for number one” and realized in acts of treating everything outside of itself as nothing but a means to this end.<sup>5</sup> One may well want to protest that Kant’s high standard for rational beings should not be deemed applicable to all living, including non- or

---

funded for the period 1 June 2020 to 31 May 2023. The project is led by Leonardo Bich and Xabier Barandiaran with Kepa Ruiz-Mirazo, Jon Umerez, and Arantza Etxeberria as members.

---

<sup>5</sup> In their 2009 paper “Defining Agency— Individuality, Normativity, Asymmetry and Spatio-temporality in Action,” (Barandiaran et al. 2009), Xabier Barandiaran, Ezequiel Di Paolo, and Marieke Rohde specifically delimit the status of agency to the subjects (versus objects) of asymmetrical acts. By 2019 (as seen above) Barandiaran had signed on to a research program in “outonomy” that seeks to transcend the limits of the autonomy paradigm.

prerational beings, and perhaps that's fair, but we have seen that Bich and colleagues, all of whom have been active in the biological autonomy research community, are specifically striving to realize, as they state in the grant proposal, a "multiscale theory of autonomy," the relevance of which includes "the emergence of collective forms of autonomous agency, the social constitution of individual autonomy through social habits, and the emergence of political autonomy in democratic societies."

The liabilities of concretizing the "*Bildungstrieb*" or any other heuristic placeholder into a research program all about elucidating the mechanistic bells and whistles of organizational closure also includes contradictions closer to home. As we've already seen, Kant's heuristic notion of a "generic preformationism" (built of *Keine und Anlagen*) could support in principle an adaptive change of structure (so long as it is within the confines of the generic scope of adaptive potential already taken as heuristically given). However, to attempt to both mechanistically cash out "autonomy" in terms of a concatenation of mechanisms that realize organizational closure and yet to also recur to a sense of a quasi-transcendent capacity for self-norming that can "shift to different and new organizational regimes," as V&M express it, is surely to want to have it both ways.

We had earlier introduced into this conversation the concept of cooperativity. We can of course allow for the meaning of autonomy as "self-norming" to refer to the well-rehearsed circular causality of the cell/organism as a regime of internal cooperativity. But why should cooperativity stop there? Are there cells and/or organisms in our world that engage in no external relations of cooperativity? I would suggest that research literature now appears on a daily basis that would contradict any such claim.<sup>6</sup> Is there a universe of meaning in which cooperative engagement would not count as a kind of agency? Not that I could fathom.

And yet if agency is held to be contingent upon autonomy and autonomy is identified as restricted to the acts of "looking out for number one," we would seem to have a problem. Surely any cell (and beyond) that engages in cooperative interaction is participating in acts of mutual or collective

<sup>6</sup> Just to evidence one very recent and fascinating example of this, on 15 June 2023 Nature Communications published a paper entitled "Bacteria Evolve Macroscopic Multicellularity by the Genetic Assimilation of Phenotypically Plastic Cell Clustering" by Yashraj Chavhan, Sutirth Dey, and Peter A. Lind. The article shows that both Gram-positive and Gram-negative bacteria have the capacity to respond to environmental changes with an indigenous capacity to form non-differentiated multicellular clusters. Recognizing that cells on the periphery of the cluster experience a different microenvironment than those at the core, the authors suggest that, "an exciting new line of work would be to test if such ecological differences can drive the evolution of cellular differentiation" (Chavhan et al. 2023). I think that it would be fair to say that an openness to collective cooperative norm formation is coming into focus.

self-norming, i.e., autonomy. We thus appear to have come to a fork in the road. We do not yet have a concrete understanding of how cells/organisms partake of norm formation. The idea of organic norm formation, i.e., autonomy, thus has the status of a reflective concept (it is what we need to make sense of what we see in the absence of a pre-given determinative law). We can go down the road of attempting to harden it into a principle of organizational closure that may allow us to begin to elucidate concatenations of mechanisms that secure a kind of internal continuity, but we have also seen what difficulties/aporias this path leads to. Alternatively, we can retain a more capacious notion of autonomy as a kind of intervening concept that holds a place for a new determinative principle which we have yet to construe and that offers a pathway that allows both mechanistic and teleological judgments to interplay. Recalling Pickering's ontology of ubiquitous agency and given the added propensity for incipient forms of cooperativity, might we not begin to see the living cell as both an organic island of internal normative stability and adaptability (as organizational closure aspired to account for) but also an interactor with aptitudes for cooperatively bootstrapping up into higher levels of coordination and becoming?

**Funding** The author acknowledges the financial support of the John Templeton Foundation, United States (#62220). The opinions expressed in this paper are those of the author and not those of the John Templeton Foundation.

## Declarations

**Competing Interests** The author has no relevant conflicts of interest to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Barandiaran X, Di Paolo E, Rohde M (2009) Defining agency—individuality, normativity, asymmetry and spatio-temporality in action. *Adapt Behav* 17(5):367–386
- Chavhan Y, Dey S, Lind PA (2023) Bacteria evolve macroscopic multicellularity by the genetic assimilation of phenotypically plastic cell clustering. *Nat Commun* 14:3555. <https://doi.org/10.1038/s41467-023-39320-9>

- Dennett DC (1989) *The intentional stance*. MIT Press, Cambridge
- Ho MW (2012) *Living rainbow H2O*. World Scientific Publishing, Singapore
- Kant I (2000) [1790]. In: Guyer P, Matthews E (trans) (eds) *Critique of the Power of Judgment*. Cambridge University Press, Cambridge
- Maturana HR, Varela FJ (1980) *Autopoiesis and cognition—the realization of the living*. Boston Studies in the Philosophy and History of Science, vol 42. Reidel, Dordrecht
- Moreno A, Mossio M (2015) *Biological autonomy. A philosophical and theoretical enquiry*. Springer, Dordrecht
- Moss L (2017) *Detachment theory: agency, nature and the normative nihilism of new materialism*. In: Ellenzeig S, Zammito J (eds) *The new politics of materialism*. Routledge, New York
- Okasha S (2023) The concept of agent in biology: motivations and meanings. *Biol Theory*. <https://doi.org/10.1007/s13752-023-00439-z>
- Pickering A (2023) What is agency? A view from science studies and cybernetics. *Biol Theory*. <https://doi.org/10.1007/s13752-023-00437-1>
- Pollack GH (2013) *The fourth phase of water*. Ebner and Sons, Seattle
- Prigogine I, Stengers I (1984) *Order out of chaos*. Bantam, Toronto
- Snow CP (1959) *The two cultures and the scientific revolution*. Oxford University Press, Oxford
- Virenque L, Mossio M (2023) What is agency? A view from autonomy theory. *Biol Theory*. <https://doi.org/10.1007/s13752-023-00441-5>
- Winfree AT (1984) The prehistory of the Belousov-Zhabotinsky oscillator. *J Chem Educ* 61(8):661–663

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.