



Concepts of Agency: Introduction to the Thematic Section

Lenny Moss¹

Accepted: 18 December 2023 / Published online: 4 January 2024
© Konrad Lorenz Institute for Evolution and Cognition Research 2024

Agency in nature has become a hot topic across a number of disciplinary boundaries. Recent and current theorizations and debates, however, have been partitioned into two non-communicating camps. On the one side, driven largely by a regard for the apparent inherent “purposiveness” of the living organism and against genetic reductionism, inspired by Kant, the autopoietic theory of Maturana and Varela (1980), and early 20th-century organicism (Nicholson 2014; Gilbert and Sarkar 2000), theorists and philosophers of biology have been grappling with how to *expand* the franchise of human agency to include some or all living organisms, albeit within a naturalist framework. Drawing upon presumptions about human agency, normativity is taken to be an irreducible aspect of any form of natural agency even while reconciling this with the demands of a fully naturalist account can be a work in progress. On the other side of the divide, in a spirit of *post*-humanism, with French philosophers Gilles Deleuze and Bruno Latour as leading lights, researchers have sought rather to *explode* the human monopoly on agency, resulting in its general dispersal throughout the material universe.¹ All matter is thereby deemed to have agentive status, albeit with questions of normativity largely sidelined if not eliminated.

The following four essays, by Okasha, Pickering, Virenque and Mossio, and Watson were invited as contributions to a special forum on “agency” with several desiderata in mind. While the principal focus of this journal is on the biological side of things, some consideration was given to the possible benefits of bridge building across this disciplinary divide. A variety of perspectives were deliberately elicited in the hope of eschewing insularity, provoking and promoting critical exchange, and perhaps even effecting

some adventitious cross-pollination. In what follows the four essays will be introduced and some possible relations between them suggested and explored. On a first pass, the essays by Okasha and Virenque and Mossio will be coupled, as will those of Pickering and Watson, on the basis of the prominent place of the concept of “autonomy” for the former and the absence of such for the latter.

Samir Okasha’s essay, “The Concept of the Agent in Biology: Motivations and Meanings,”² is unique amongst the group inasmuch as rather than putting forward its own position on agency, Okasha sets out to analytically clarify the use and understanding of “agency,” at least on the biology side of the fence. As his title suggests, Okasha takes a very straightforward approach in first asking why anybody would be interested in invoking the concept of agency, i.e., the motivation for using it, and then proceeding to line up alternative senses of “agency” drawn from various external (to biology) disciplinary uses in order to find the best match between motivation and meaning. Okasha gestures in the direction of looking to different disciplinary employments of the concept of “agency” as a safeguard against biology simply declaring a class of entities as agentive by definitional fiat. Okasha’s essay may well in fact highlight this issue as worthy of further consideration.

With respect to motivations, these appear to divide cleanly into realist and anti-realist perspectives on agency in nature. The realist motivation rejects genetic reductionism and perceives organisms as the irreducible locus of various types of actions and attributes indicative of agency. Whether the status of “autonomy” confers a unique sense of agency upon the organism, as opposed to any other level of biological organization, is a topic of debate within the realist camp. The anti-realist sense, by contrast, does not recognize the organism as a bona fide agent but rather sees some heuristic benefit in speaking “as-if” the organism possessed intentional attitudes such as beliefs and desire, when in fact it is the forces of natural selection (mechanistically mediated by genes and genetic programs) that are pulling the strings.

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

✉ Lenny Moss
dcimorse@icloud.com

¹ Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México, Mexico City, Mexico

² This issue; <https://doi.org/10.1007/s13752-023-00439-z>.

Okasha approaches his battery of agency concepts through first establishing a minimalist and a maximalist concept at the antipodes of his line-up. The minimal would be just any source of action in which the motive force is internal as opposed to external, while the maximal concept would be one in which an activity is the expression of a prior mental intention. Intermediate concepts would entail some form of normativity, absent from the minimalist concept, but without the higher requirement of some form of prior mental representation. As intermediate concepts, Okasha identifies an “intelligent agent” concept from artificial intelligence (AI) and a “rational agent” concept from economics. He goes on to surmise that the intelligent agent concept from AI, which specifies a capacity for flexible responses in relation to environmental contingencies, would be the best fit for a realist theory of biological agency. The fit, however, is not perfect insofar as it does not meet the expectation of picking out what is unique to organisms, i.e., “autonomy,” as opposed to other biological entities either beneath or above the level of the organism. This result leads nicely into the essay by Virenque and Mossio that explicitly aims to elaborate upon an “autonomy”-based theorization of biological agency.

In their essay, “What is Agency? A View from Autonomy Theory,”³ Louis Virenque and Matteo Mossio refer to the language (if not the determinative versus reflective epistemic distinction) of Kant as their point of departure in characterizing the meaning of an autonomous system. An autonomous system is that which is a “cause and effect of itself,” which is to say that it is composed and organized in such a fashion that its components produce that system that in turn reproduces its components and self-producing organization. Inasmuch as the component parts must be replaced over time and the system must maintain a steady input of energy in order to maintain its nonequilibrium regime, the system must achieve an organizational closure that enables it to both exchange matter and energy with its environment and yet conserve and secure its organizational integrity and function. In so doing the organism constitutes a form of agency that is arguably limited to organisms. Virenque and Mossio tell us that such a conception of agency meets three distinctive requirements (which we will see are necessary but not sufficient): that of being non-intentional (i.e., not meeting the requirements of Okasha’s strong concept), intrinsic (as the minimal concept already requires), and naturalized (by which we can understand as meaning empirically warranted). With respect to this latter criterion, the justification of which would clearly exceed the limits of a single position essay, they acknowledge it to be “a point that we take for granted here.”

In the light of challenges to the tradition from which their concept of an autonomy-based notion of agency was derived, they recognize that a requisite notion of agency must be accountable to challenges with respect to adaptability and from adaptability to a capacity for “sense-making” of the environment to which they must adapt. Adaptability, however, implies that the autonomous system be able to realize its self-purposiveness at a more general level than that of rigidly preserving its organizational structure, but rather must be able to adapt its organization. The remit of self-purposiveness must be more general than that constituted by any invariant organization. They further complexify their treatment of agency by looking ahead to the possible tension between a purely practical propensity for environmental sense-making and the onset of cognition as an agentic attribute that holds itself accountable to epistemic and not just survivalist standards. These “complexifications” are best understood as programmatic challenges for the further elaboration of the theory.

The final two essays of this collection, those of Pickering and Watson, do not turn on concepts of autonomy (and the traditions from which the idea of biological autonomy arose) nor necessarily wish to treat life, let alone the specificity of the organism, as thresholds, *sine qua non*, for the identification of natural agency. Indeed with Pickering we encounter the perspective of the “ontological turn” in science studies also seen in the work of many across the humanities who have taken a decidedly post-humanist turn. In his essay, “What is Agency? A View from Science Studies and Cybernetics,”⁴ Andrew Pickering finds the experimental physicist, drawing on his own well-known paradigm example of quark detection (Pickering 1984), as engaged in an ongoing process of adjustments and recalculations, a veritable “mangle of practice” Pickering later prefers to call a “dance of agency,” that continues until “an island of stability” is discovered. Only in the context of such an island of stability, whereby the dance of agency is held in abeyance, does the agency of nature become “backgrounded” and the familiar relation of human subject to natural object become envisaged. Materiality, nature in general, are as primordial oceans of active and lively agency, at least in the minimalist sense of Okasha. Pace Virenque and Mossio (and much of the current associated literature), agency neither presupposes nor begins with agents for Pickering. Rather, agents are transient beings that congeal from out of fields of agency. Might it be the case then that Pickering is adumbrating a pathway for dissolving the demarcation between life and nonlife, but if so how are we to reckon with the emergence of normativity from nonnormative minimal agency? In approaching questions of normativity (if not in so many

³ This issue; <https://doi.org/10.1007/s13752-023-00441-5>.

⁴ This issue; <https://doi.org/10.1007/s13752-023-00437-1>.

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 still 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. This then leads us nicely to the final essay.

Although principally addressed to the topic of biological agency, Richard Watson's essay, "Agency, Goal-Directed Behavior, and Part-Whole Relationships in Biological Systems,"⁵ foreswears any assumption of biological givens in its approach to a naturalist account of agency: "we avoid trying to identify any vitalist criterion that categorically distinguishes between agential and non-agential ('ordinary mechanistic') systems in absolute terms." Indeed the logic of the argument is largely spelled out using non-biotic exemplars such as balls rolling down an incline. Beginning with a minimalist depiction of agency (or none at all) and proposing the possibility of a spontaneous emergence of an agentive system, we will want to consider whether Watson has collaterally provided Pickering with a pathway toward normative agency and perhaps even a conceptual bridge to the strongly normative account of Virenque and Mossio.

The crux of Watson's approach to agency is to ask whether a system has acquired the ability, as a system, to enable its parts to achieve their goals better than they would independently. It is thus an approach to agency based on distinguishing parts from the whole of a system. Owing to the absence of "vitalist criteria" the idea of a part having a goal can be modelled as simply as a ball rolling down an incline and resting at local minima. In a landscape in which the achievement of even greater energy minima could be had but only if the relevant part could resist the local minima and find its way, bypassing immediate "gratification," to the greater goal, can one then speak of the latter situation as more agentive than the former. A system of parts that, as a system, and through experience, undergoes modifications of the spatial and/or temporal ordering of its parts such that the parts can better overcome constraints and achieve higher goals becomes the model for natural agency (even if just in relative terms). Watson offers a baseline conception of the emergence of a higher-order relative agency drawing upon a Hebbian feedback model:

It has been shown that this modification can be provided by unsupervised Hebbian learning—but requires no external teacher or reinforcement signal—but simply applies a positive feedback on observed correlations. . . Specifically, the system is allowed to spend most of its time at local attractors, with occasional shocks that cause it to visit a distribution of attractors over time. Over a timescale where many such attractors are visited, the weights of the network are slightly modified by Hebbian learning, and the network slowly learns an associative model of the configuration it visits.

Watson gives us a tempting picture of a possible emergence of a simple normativity based upon physicalist attractors, but we may wonder whether there is at least theoretically a pathway to the stronger normativity, as Virenque and Mossio recount, of the circularity of parts striving to reproduce themselves through reproducing each other? Can the transition into life be reckoned as a higher-order attractor? Less dramatically, does Watson offer us an optics for construing the pervasive minimal agency (*sensu* Okasha) of Pickering as susceptible of gaining fledgling normativity by way of an incipient systematic cooperativity of the parts?

Funding and Competing Interests The author has no relevant financial or nonfinancial conflicts of interest to disclose.

References

- Gilbert S, Sarkar S (2000) Embracing complexity: organicism for the 21st century. *Dev Dyn* 219:1–9
- Maturana HR, Varela FJ (1980) *Autopoiesis and cognition: the realization of the living*. Boston studies in the philosophy of science, vol 42. Reidel, Dordrecht
- Moss L (2017) Detachment theory: agency, nature, and the normative nihilism of new materialism. In: Ellenzweig S, Zammito J (eds) *The new politics of materialism*. Routledge, New York, pp 227–249
- Nicholson DJ (2014) The return of the organism as a fundamental explanatory concept in biology. *Philosophical Compass* 95:347–359
- Pickering A (1984) *Constructing quarks: a sociological history of particle physics*. University of Chicago Press, Chicago

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

⁵ This issue; <https://doi.org/10.1007/s13752-023-00447-z>.