What is Agency? A View from Autonomy Theory

To appear in *Biological Theory*

The final publication is available at Springer via:

https://doi.org/10.1007/s13752-023-00441-5

Louis Virenque[1]

ORCID : 0009-0007-1194-716X

Matteo Mossio[1]*

ORCID : 0000-0003-0831-0815

[1]IHPST, CNRS/Université Paris 1, Paris, France

*corresponding author: Matteo.Mossio@univ-paris1.fr

Abstract

The theory of biological autonomy provides a naturalized characterization of agency, understood as a general biological phenomenon that extends beyond the domain of intentionality and causation by mental states. Agency refers to the capacity of autonomous living beings (roughly speaking: organisms) to purposively and functionally control the interactions with the environment, and to adaptively modulate their own self-determining organization and behavior so as to maintain their own existence, construed as their intrinsic telos. We mention some crucial strengths of the autonomist conception of agency, and some interesting challenges that it faces. Among the latter, we focus on the intertwined relationships between agency and evolution, as well as on the transition between agency and cognition.

Keywords

Adaptivity; Agency; Autonomy; Autopoiesis; Cognition; Evolution; Purposiveness

Introduction

In the mainstream conception, the notion of agency is related to intentionality. Action is intentional behavior, which means behavior performed for a reason, oriented toward a goal. A behavior, in turn, can be said to be performed for a reason only if it is caused by certain mental states (as desires and beliefs) that have a representational content related to the goal and the means to attain it (Schlosser 2019). In such a conception, agency is usually attributed to a very specific class of living systems, namely human beings (Frankfurt 1978; Davidson 1982). Yet linking agency to the mind is not the only possible stance; living beings at large can also be characterized as agents by relying on a more general (and yet naturalized) understanding of purposiveness. The theory of biological autonomy provides such a characterization. In a nutshell, the theory of biological autonomy holds that a system is an agent if it is capable of interacting with its environment in such a way that its behavior is, first, enabled by its own constitutive organization and, second, contributing to maintain that very organization (Barandiaran et al. 2009; Arnellos and Moreno 2015; Moreno and Mossio 2015).

The origins of the theory of autonomy can be traced back to Kant ([1790]1987) and, more recently, to Varela (1979). A central leitmotif of older and more recent accounts of autonomy is the idea that it is not possible to adequately make sense of the nature and behavior of a living being by appealing only to mechanistic methods and concepts, which consist of explaining a phenomenon in terms of the properties of—and interactions among—the constituents of the relevant system. Instead, 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." Thereby, living beings are (in Kant's terminology) *self-organizing* natural systems. In particular, and in contrast to mechanistic systems, the constituents of self-organizing systems at the same time produce and are produced by the totality to which they belong. As a consequence, the explanatory

relationship becomes circular: the properties of (and interactions among) the constituents account for the whole organization, and vice versa.

By relying on the circular organization of living beings, the theory of autonomy provides a naturalized ground for several concepts whose scientific legitimacy beyond the human domain is questioned, such as goals, norms, function, and, in particular, agency. Agency, therefore, is accounted for by means of a specific characterization of living organization, which implies making it an inherently *biological* phenomenon. The theory of autonomy thereby attempts to bring back to the biological realm what has been neglected since the advent of the Darwinian theory of evolution by natural selection—a neglect reinforced by the Modern Synthesis during the 20th century (Walsh 2015).

Naturalizing Agency from the Perspective of Autonomy

According to the theory of autonomy, the mutual determination between the whole and its parts is the fundamental feature of a natural agent's constitutive organization, which notably differentiates living beings from artifacts and machines. In the literature to which we refer, such feature is sometimes called autopoiesis[1] (literally: self-production). Autopoiesis does not mean "self-creation" in the sense of spontaneous generation, but instead reinterprets in contemporary terms what Kant ([1790]1987) labels "formative force" in his C*ritique of Judgement*. 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,

[1] A canonical definition of autopoiesis reads as follows : "An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as a network" (Maturana and Varela 1980, pp. 78–79).

no force is at play here: the *organization* of the parts is such that they collectively contribute to their own existence.

The self-determining nature of biological organization allows referring to *purposiveness* in a legitimate way: insofar as the effects of their activity contribute (at least in part) to determining their existence, the purpose of living beings is their own organization (Mossio and Bich 2017). Relatedly, the *function* of each part is to maintain the whole organization and, thereby, to maintain itself (Mossio et al. 2009). Unlike other categories of purposive systems, such as artifacts and machines, which require an extrinsic purpose to be produced, living being are *intrinsically* purposive, given that the reason why they exist is...themselves (their own organization). Machines, in contrast, exist because they are means to achieve the goals of a third party (which might be an individual animal or person, or a more complex socio-technical community).

The mutual dependence among the functional parts of a purposive organization is usually referred to as "organizational closure" (Montévil and Mossio 2015). Organizational closure is a condition for the maintenance of the system because biological parts undergo degradation over time, and must therefore be repaired or replaced. As is often underscored in the theory of autonomy (Ruiz-Mirazo and Moreno 2004), living beings are far-from-thermodynamic equilibrium systems, and their existence requires a continuous exchange of matter and energy with the environment, so as to feed the metabolism and re-establish the organization while "locally contravening" the second law of thermodynamics.

Insofar as organizational closure implies thermodynamic nonequilibrium, living beings as autonomous systems are not autarkic or independent but, rather, they must continuously *interact* with their surroundings. The interaction itself is controlled by functional parts and subsystems subject to organizational closure: hence, interactive capacities are themselves intrinsically purposive. Such purposive, functional interactive capacity performed by the living being realizing organizational closure is *agency* within the theory of autonomy (Barandiaran et al. 2009). 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. Examples of actions performed by agents are the pursuit of a bacterium by a neutrophil, the phototropism of a plant, and the foraging of a rabbit, insofar as they are interactive behaviors that contribute to the maintenance of the organization, which enables them. The theory of autonomy offers, therefore, a perspective from which agency can be understood as behavior performed for a reason, directed towards an intrinsic goal, which is the continued existence of the system's self-determining organization, through an incessant interaction with its external environment.

The distinctive features of such a conception of agency are therefore threefold: non-intentional, intrinsic, and naturalized. First, agency is not necessarily related to intentionality and the mind, even though human-specific agency can, at least to some extent (but see below), be construed as a special case of natural agency. Second, the purposiveness of agency is intrinsic, stemming from the organizational closure of living beings; it should then be distinguished from the extrinsic purposiveness of artifacts, which depends on the reference to an external designer or constructor. In this respect, a crucial contribution of the theory of evolution by natural selection has been to provide a scientific alternative to a teleological explanation of biological phenomena appealing to a divine creator. Yet the theory of autonomy emphasizes the importance of the distinction between extrinsic and intrinsic purposiveness for biological and cognitive science: abandoning the first should not prevent the central role of the second from being acknowledged, especially with regard to the concept of agency. Third, while being associated to neither intentional nor extrinsic purposiveness, 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 standards of natural science (a point that we take for granted here).

Construed as the interactive dimension of organizational closure, agency includes all behaviors performed towards the whole living being's overarching *telos:* its own preservation as a far-from-equilibrium organized system. Whatever their specific effect is, all functional organs and parts (and specifically those performing actions), are supposed to contribute to the intrinsic purpose

of the organism as a whole. As we discuss below, however, this general stance requires qualification, because such a minimal agent is not yet an autonomous system, the concept that the theory employs to characterize a living being (which is, to a first approximation, an organism).[2] The idea of biological autonomy calls for a more sophisticated conception of agency.

Complexifying the Agent

If we were to stop at the characterization of agency given above, we would expose ourselves to the now classical criticism addressed by Di Paolo (2005) to the definition of autopoiesis by Varela and Maturana (1980, see footnote 1). According to Di Paolo, a pure autopoietic system is able to survive in a particular, stable environment by relying on its self-determining organization, but it is unable to adapt to changing conditions, a capacity that Di Paolo refers to as *adaptivity*. An adaptive system is a system that is able to undergo functional modifications so as to deal with internal or external perturbations.[3] In turn, an adaptive system is an adaptive agent if such modifications specifically affect its interactive capacities. Compared to minimal agency, adaptive agency involves more sophisticated skills, including higher-order regulation and anticipation, as well as the possibility to shift to different and new organizational regimes. As claimed by Moreno and Mossio (2015, Chap. 4), an organizationally closed adaptive agent is an autonomous system and, thereby, a living system. As a matter of fact all living systems, be they unicellular or multicellular, meet ex hypothesi the

[2]While minimal agency appears to be necessary but not sufficient to characterize autonomy (and organismality), not every biological system is necessarily an agent. For instance, an ecosystem's organization might possibly be shown to realize closure and, thereby, be considered as a biological system (Nunes Neto et al. 2014); yet this would not necessarily imply that the ecosystem is also an agent. We do not address these questions here, but it is important to keep in mind that concepts such as closure, agency, and autonomy are not only conceptually distinct, but they could also apply differently to various empirical cases.

[3]Di Paolo's definition of adaptivity reads: "A system's capacity, in some circumstances, to regulate its states and its relation to the environment with the result that, if the states are sufficiently close to the boundary of viability, tendencies are distinguished and acted upon depending on whether the states will approach or recede from the boundary and, as a consequence, tendencies of the first kind are moved closer to or transformed into tendencies of the second and so future states are prevented from reaching the boundary with an outward velocity" (2005, p. 438).

characterization in terms of adaptive agents (including the examples given above). [4] 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.

One important implication of adaptive agency is that its realization leads to what is referred to as "sense-making" (Weber and Varela 2002, p. 18), i.e., the fact that the agent makes sense of its environment in relation to its intrinsic purposiveness. To mention a classical Varelian example,

> That sucrose is a nutrient isn't intrinsic to the structure of the sucrose molecule; it's a relational feature, linked to the bacterium's metabolism. Sucrose has significance or value as food, but only in the milieu that the organism itself brings into existence. (Thompson 2004, p. 286)

Thompson (2007) parallels Varela's sense-making with Uexküll's notion of the *Umwelt* (1934), elaborated in the context of his work on the perception of their environment by (human and nonhuman) animals. Another crucial implication is that an adaptive agent must be able to *sense* the environment, so as to detect changes and perturbations with respect to which an appropriate action is performed (Moreno 2018).

The ability to make sense of one's environment can be understood as one of the first steps toward more complex forms of agency and cognition. In the theory of autonomy, the question whether there is a difference between agency and cognition is the object of an ongoing debate, notably within enactivism (Bourgine and Stewart 2004; Di Paolo et al. 2017; see also Gambarotto and Mossio 2022). According to one particular position, cognition is qualitatively different from agency insofar as it designates behavioral and interactive capacities whose purpose goes beyond the fundamental one, i.e., the preservation of its own existence. One can deal with this issue from an evolutionary perspective

[4] There is a debate within the theory of autonomy about whether, insofar as virtually all existing living systems are adaptive agents, only adaptive agency should count as genuine agency (see Moreno 2018 for a discussion). Here, we do not take a position on this debate, and we limit ourself to noting that 1) minimal agency has the merit of pinpointing the fundamental features of the concept (notably those discussed by Barandiaran et al. 2009) and 2) it may be that minimal and adaptive agency can be separated empirically, for example in the context of investigations into the origins of life.

and argue that, starting from the general sense-making capacity of agents, more complex skills have emerged little by little, up to the realization of the cognitive systems that we know today. In addition, the theory of autonomy also looks at the relationship between agency and evolution the other way around, by exploring how agents shape evolution through their reciprocal interactions and their influence on the environment (Walsh 2015; Sultan et al. 2022). Such a perspective on the complex relations between agency, cognition, and evolution participates in a trend, which has brought into the spotlight phenomena that seemed to be underestimated by the Modern Synthesis, such as niche construction and developmental plasticity, and which puts the organism as an agent at the center stage (Lewontin 1985; West-Eberhard 2003; Bateson 2005; Laland et al. 2014; Sultan 2015). The theory of autonomy makes an original contribution to structuring this trend, thanks to the organizational rooting of biological agency and cognition on which it relies.

Yet the evolutionary continuity between agency and cognition should not overlook their organizational discontinuity. As mentioned, the theory of autonomy grounds the purposiveness of adaptive agency, enhanced with sense-making, in terms of the contribution to the intrinsic telos, which is an organized system's own existence. Given that intrinsic purposiveness is *by definition* construed as a circular relation between the existence and the activity of a system, it follows that any function or action performed by an autonomous system is purposive *insofar as* it contributes to determining its conditions of existence. However, the philosophical problem raised by cognition is that there seems indeed to exist a kind of purposive behavior that, prima facie, does not contribute to an agent's own survival. Different strategies can be envisioned to deal with this issue, and we do not discuss them here.[5] The main philosophical choice seems to consist in either arguing that purposiveness does not need to be anchored to an intrinsic telos, or that any purposeful behavior can be shown, in fine, to contribute to the existence of a self-determining system. Whatever the choice, what is at stake is an adequate understanding of the connection and difference between *surviving* and *living.*

[5]As a matter of fact, the same kind of problem applies to reproduction, which seems also to be a biological phenomenon in which purposeful behavior does not contribute to the preservation of the agent itself. Advocates of the theory of autonomy have dealt with reproduction in previous publications (see Saborido et al. 2011; Mossio and Pontarotti 2019).

To conclude, the theory of autonomy provides an understanding of agency as a biological phenomenon, grounded in the self-determining purposeful organization of living beings. In particular, agency designates the functional capacities devoted to governing the interaction of the organism with the external environment (which of course includes other living beings). Agency, and in particular adaptive agency, is one of the central dimensions of the overarching idea of biological autonomy. The understanding of agency from the perspective of autonomy possesses some crucial strengths and faces interesting challenges. Among the latter, we have mentioned the complex relationships between agency and evolution, as well those between agency and cognition. We look forward to seeing how the theory of autonomy will take up such challenges in the future. In particular, the project of elaborating an account of the transition between agency and cognition requires dealing with the role played by the nervous system and the brain in enabling the emergence of more sophisticated purposive behavior (Barandiaran and Moreno 2006). In turn, this opens the way to a biologically grounded account of the mind (Thompson 2007; Di Paolo et al. 2017) and, possibly, to establishing a connection with the conceptions of agency appealing to intentionality and mental states.

Declarations

Competing interests

The authors have no relevant financial or nonfinancial interests to disclose.

Funding

Funding for this research was provided by the CNRS—University of Toronto "PhD Mobility Joint Program" (PhD Fellowship to Louis Virenque).

References

Arnellos, A. and Moreno, A. (2015). “Multicellular agency: an organizational view”. Biology and Philosophy 30(3): 333-357.

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

Barandiaran, X., & Moreno, A. (2006). On what makes certain dynamical systems cognitive. Adaptive Behavior, 14, 171–185.

Bateson, P. (2005) “The return of the whole organism”. *J Biosci* **30**, 31–39 (2005).

Bourgine, P., Stewart, J. (2004). “Autopoiesis and cognition”. *Artificial Life*, 10, 327-346.

Davidson, D. (1982). “Rational Animals”, *Dialectica*, 3(4): 317–327.

Di Paolo, E. A. (2005). “Autopoiesis, adaptivity, teleology, agency”. *Phenomenology and the Cognitive Sciences*, 4(4), 429–452.

Di Paolo, E., Buhrmann, and Barandiaran, X. (2017). *Sensorimotor Life: An Enactive Proposal*, Oxford: Oxford University Press.

Frankfurt, Harry G. (1978). “The Problem of Action”. *American Philosophical Quarterly* 15 (2):157-162.

Gambarotto, A., Mossio, M. “Enactivism and the Hegelian Stance on Intrinsic Purposiveness”. *Phenom Cogn Sci* (2022).

Kant, E. ([1790]1987). *Critique of judgment*. Indianapolis: Hackett Publishing.

Laland K, Odling-Smee J, Endler J. (2017) “Niche construction, sources of selection and trait coevolution”. *Interface Focus* 7: 20160147. http://dx.doi.org/10.1098/rsfs.2016.0147

Lewontin, R. C. (1985). “The Organism as Subject and Object of Evolution”. In R. Levins and R. Lewontin, *The Dialectical Biologist* (pp. 85–106). Cambridge: Harvard University Press.

Maturana, H., and Varela, F. (1980). *Autopoiesis and cognition. The realization of the living*. Dordrecht: Reidel Publishing.

Montévil, M., & Mossio, M. (2015). “Biological organisation as closure of constraints”. *Journal of*

Theoretical Biology, 372, 179–191.

Moreno, A., “On Minimal Autonomous Agency: Natural and Artificial”. *Complex Systems*, vol 27, 3, 289-313 (2018).

Moreno, A., Mossio, M. (2015). *Biological Autonomy. A Philosophical and Theoretical Enquiry.* Dordrecht, Springer.

Moreno, A., Ruiz-Mirazo, K. “The problem of the emergence of functional diversity in prebiotic evolution”. *Biol Philos* **24**, 585–605 (2009).

Mossio, M., Bich, L. (2017). “What makes biological organisation teleological?” *Synthese*, 194, 4, 1089-1114.

Mossio M, Pontarotti G (2019) Conserving functions across generations: heredity in the light of biological organisation. *British Journal for the Philosophy of Science*, 73, 1, 249-278.

Mossio, M., Saborido, C., Moreno, A. (2009), “An organisational account of biological functions”. *British Journal for the Philosophy of Science*, 60, 4, 813-841.

Nunes, N., Moreno, A., El Hani, C. (2014). “Function in ecology: an organizational approach”. *Biology and Philosophy*, 29, 1, 123-141.

Ruiz-Mirazo, K., & Moreno, A. (2004). “Basic autonomy as a fundamental step in the synthesis of life”. *Artificial life*, *10*(3), 235–259.

Saborido, C., Mossio, M., & Moreno, A. (2011). Biological organization and cross-generation functions. The British Journal for the Philosophy of Science, 62, 583–606.

Schlosser M (2019) Agency. In: Zalta EN (ed) The Stanford encyclopedia of philosophy (winter 2019 edn). https://plato.stanford.edu/archives/win2019/entries/agency/

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

Sultan SE, Moczek AP, Walsh D. (2022) “Bridging the explanatory gaps: What can we learn from a biological agency perspective?” *Bioessays* 44(1):e2100185

Thompson, E “Life and Mind: From Autopoiesis to Neurophenomenology. A Tribute to Francisco Varela.” *Phenomenology and cognitive sciences* 3.4 (2004): 381–398.

Thompson, E. (2007). *Mind in Life. Biology, Phenomenology, and the Sciences of Mind*. Cambridge, Harvard University Press.

Uexküll, J. von (1934/2010). *A Foray Into the Worlds of Animals and Humans: With a Theory of Meaning*, Minneapolis/London, University of Minnesota Press.

Varela, F. J. (1979). *Principles of biological autonomy*. New York: North Holland.

Walsh, D. (2015). *Organisms, Agency, and Evolution*, Cambridge, Cambridge University Press.

Weber, A. Varela, F. (2002). "Life after Kant: Natural purposes and the autopoietic foundations of biological individuality". *Phenomenology and the Cognitive Sciences*, 1, 97- 125.

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