

I ACT THEREFORE I LIVE? AUTOPOIESIS, SENSORIMOTOR AUTONOMY, AND EXTENDED AGENCY

– Barbara Tomczyk –

Abstract: This paper aims to determine whether extended human-machine cognitive systems and group systems can be regarded as autonomous agents. For this purpose, I compare two notions of agency: one developed within analytical philosophy of action and based on the concept of intention, and the other introduced by enactivists via the concepts of autopoiesis and sensorimotor autonomy. I argue that only the latter approach can be used to demonstrate autonomous agency in respect of systems that are not humans as such, though they contain humans as their elements. After introducing Maturana and Varela's conception of minimal autonomy as a kind of generalization of autopoiesis, I present the three conditions of agency put forward by Barandiaran, Di Paolo and Rohde, noting that they do not invoke the property of being alive as necessary in that respect. I argue that both extended and group systems can satisfy these conditions of agency, even though they are not alive as such. The fulfillment of these conditions, however, is ensured by the autopoietic nature of the living components of these systems. That being said, an autonomous system itself does not need to be alive in the biological sense. Sensorimotor, adaptive agency could emerge out of other processes than those responsible for biological life. The article concludes with a suggestion that this is exactly what will happen if an autonomous system is ever artificially created. It would be functionally indistinguishable from a living organism, though not alive in a biological sense.

Key words: autopoiesis, agency, extended cognitive system, group system, sensorimotor autonomy, enactivism

Submitted: 26 October 2022

Accepted: 18 February 2024

Published online: 11 May 2024

1. Introduction

Now that roboticists and programmers are designing systems that solve complex cognitive tasks in ways increasingly independent of humans, the concept of agency has become central to cognitive science and artificial intelligence research. Such research programs are motivated by an aspiration to create artificial systems, such as sensorim-

Barbara Tomczyk
Chair of Logic and Cognitive Science
Maria Curie-Skłodowska University in Lublin
email: barbara.tomczyk@mail.umcs.pl

otor robots and deep learning-based algorithms, that could be considered autonomous agents. However, the concept of agency used by scientists involved in this venture differs greatly from the definition developed by analytical philosophers of action. According to the best-known representative of the latter, Donald Davidson, an agent must be in possession of mental states such as beliefs, desires and intentions, which are propositional attitudes. Without the employment and understanding of conceptual language, these conditions cannot be met, and the only cognitive systems that possesses this capacity are, as Davidson argues, human beings.¹ Moreover, the majority of philosophers seeking to define the conditions for autonomous action identify its subject with human beings: autonomy, on their view, is essentially related to the possibility of reasoning, awareness and evaluation as regards motives and reasons for a given action, as well as the ability to make choices between various activities, and the feeling of agency. Only then can the system be held responsible for a given action, and so be given credit or receive blame for it. This is the essence of being a full-blooded agent.²

Such an understanding of agency would appear insufficient in the age of extended cognitive systems and cognitive artifacts that are not humans as such. An alternative conception of agents is called for to properly designate cognitive systems that do not meet the conditions for assigning them full rights and obligations. To this end, cognitive scientists and AI researchers have developed the notion of an agent as a relatively autonomous system that processes information and operates in a given environment for a purpose specified by its designer. Autonomy is understood here as the ability of the system to independently change its states and the rules that govern these changes.³ Yet representatives of such a view seldom trouble themselves with specifying how this “independence” is to be understood. For that to be accomplished, the focus should be on examining different cognitive systems with a view to identifying the causes and purposes of their activity – the approach taken by those researchers who define autonomy with reference to autopoiesis as the minimum condition for agency. Although the autopoiesis requirement may seem too demanding for non-living cognitive systems, it does effectively capture a common intuition of what needs to be in place for them to be considered autonomous agents – an intuition referring to the properties of self-production (self-distinction) and self-maintenance. So, is living itself a necessary condition for being autonomous, or not? The aim of this paper is to evaluate whether extended cognitive systems (systems composed of a human and a cognitive artifact), and specific groups, could fulfill the conditions for autonomous agency as presented by the philosophers who seek to define it. The question of whether an entirely non-human, artificial system can exhibit the necessary features of autonomy is beyond the scope of this article, as it requires separate treatment.⁴ Nevertheless, analyses concerning the autonomous agency of extended and group systems are the first step to identifying the obstacles that

¹ Davidson (1982).

² Frankfurt (1971), Dworkin (1988), Chisholm (1976), Markosian (1999).

³ Chopra, White (2011), Floridi, Sanders (2004). An ultra-minimalist definition of agency can be found, for example, in the work of Stuart Russell and Peter Norvig, who write that “an agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors.” Russell, Norvig (1995): 33.

⁴ For this discussion, see Froese, Ziemke (2009).

lie in the way of attributing intentional agency to cognitive artifacts and the respective conditions that these systems have to meet.

2. Autopoietic Systems As Minimally Autonomous

According to analytic philosophy of action, as represented by Donald Davidson, agency is ascribed to a system that can make a decision to act based on the reasons it possesses. This does not mean that every action must be preceded by the agent's weighing up of reasons: it is sufficient that were he or she to be aware of them, he or she would be able to justify the rationality of his or her action on their basis.⁵ To meet these requirements, a cognitive system must have mental states with representational content, and for this a grasp of conceptual language is needed. As such, to be an agent, one has to think: thinking and acting are mutually dependent. And if one thinks, then it surely ought to be the case that one is also a participant in the exchange of linguistically conveyed meanings with other people, knowing the truth conditions of the sentences he or she utters. Thinking is an action of a specific sort in some given circumstances that consist in expressing an appropriate attitude – such as an assertion or doubt, or the wish that something be true, etc. – towards some content. An agent must therefore be in possession of the concepts of truth and falsity: otherwise, he or she would not be able to adopt any attitude towards the content in question nor undertake its normative evaluation.⁶ Consequently, on this position, propositional attitudes are indispensable for agency. There are, however, plenty of cognitive systems that do not operate using propositional language, even though they “act” in a way such that we are inclined to ascribe to them some kind of autonomous agency. Hence, in the following, I will be focusing on definitions of the agent as autonomous, though not necessarily endowed with a mind, which has been proposed by theorists of autopoiesis and then developed by enactivists.

Enactivism is an approach to cognition that draws on the one hand from phenomenological thought, and on the other from a critique of the standard, computational account of representation. According to its proponents, a cognitive agent is a being that co-creates the environment it experiences and acts in. Experiencing the world through the sensorimotor system furnishes the agent's initial situation – one that then determines all its thoughts and actions. The idea of “enaction” as a dynamic sensorimotor activity thus comes to displace the notion of “computation” central to standard cognitive science. Francisco Varela, Evan Thompson, and Eleanor Rosch, who initiated this line of thought about cognition in their book “*The Embodied Mind: Cognitive Science and Human Experience*”, point to anti-representationalism as one of the principal features of enactivism:

We propose as a name enactive to emphasize the growing conviction that cognition is not the representation of a pregiven world by a pregiven mind but is rather the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs.⁷

⁵ Davidson (1978).

⁶ Davidson (1982).

⁷ Varela et al. (1991): 9.

When speaking of cognitive processes, enactivists refer mainly to perception, which, they argue, proceeds without involving internal, symbolic representations. Such a conception of a cognitive agent implies a radical version of the embodied mind thesis, as it is through the body that an agent interacts with its environment, co-creating all of its experience. However, for enactivists it is not just perceptual experience that emerges in the interaction between the active organism and elements of its environment: even such a complex experience as the sense of one's own identity – that is, the feeling of being an autonomous agent – can be understood as a quality that emerges out of this deep-level embeddedness.⁸ Where humans are concerned, the process of self-identification and self-creation unfolds in such close coupling with the physical and social processes occurring in their environment that it becomes impossible to distinguish the internal components from the environmental components of that process. A sense of one's own agency, personal identity, or – more generally – self, is not something the agent inherently possesses, but rather emerges from being and acting with other people.⁹

Strategies for examining the properties of autonomy can be divided, just like in the case of intentionality, into those that describe a given system as autonomous on the basis of some features of its behavior, and those that seek to discern its autonomy as a function of its intrinsic properties.¹⁰ One advocate of the first of these approaches is Daniel Dennett, who introduces the strategy of the intentional stance. According to him, autonomy is assigned to the system by the observer in order to explain and predict its behavior. The question of its authentic autonomy is regarded as pointless, if not even meaningless.¹¹ Below, however, I shall be looking more closely at the solutions of researchers who consider the intentional stance strategy insufficient and seek to identify some features of the generative mechanism underlying a system's behavior that determine its autonomy in a less observer-dependent fashion, although, as I will point out in the next section, they do not perceive the system's autonomy as a completely observer-independent quality. Such a strategy is adopted by, among others, Francisco Varela and Humberto Maturana, who introduced the concept of autopoiesis in order to explain the mechanism underlying living systems.

Using the concept of autopoiesis, Varela initially focused on the description and explanation of biological processes at the cellular and meta-cellular levels of organisms; he would go on to propose the notion of autonomy as a generalization of autopoiesis that would be applicable to systems that are not biological.¹² In other words, autopoiesis is understood as a form of autonomy in the biochemical domain; however, there can be autonomous systems that are not autopoietic, namely they satisfy all of the conditions of an autopoietic entity except being biologically alive.¹³ A central idea of the theory of autopoiesis is that a living organism is not only capable of responding in a certain way

⁸ I am grateful to an anonymous reviewer for suggesting the term “emergence” for naming the relation between perceptual experience and the interaction that a perceiving organism enters into with its environment.

⁹ Kyselo (2014).

¹⁰ Rohde, Steward (2008).

¹¹ Dennett (1976).

¹² Varela (1979).

¹³ Thompson (2007).

to specific stimuli, but also has a capacity for self-constitution and self-regeneration.¹⁴ An autopoietic system is organized as a network of processes that produce components such that they continuously regenerate and realize the network that produces them, and they constitute the system as a distinguishable unity in the domain in which they exist.¹⁵ It is inherently self-producing and, as such, distinguishes itself from other entities in the environment. The minimal autonomy of such a system is determined by the feature of organizational closure: its behavior is fixed by its internal states alone, even though they may be influenced by external factors.¹⁶ In other words, the coupling of processes occurring within an autonomous system constitutes a dynamic unity that engages in interaction with the environment in order to obtain the resources necessary for its survival. Thus, environmental factors trigger changes in the structure of autopoietic systems but cannot control them. In explaining these processes, Maturana and Varela refer to the neural systems of living organisms and the changes these undergo. Their central assumption is that neural activity both stems from, and further engenders, neural processes in a closed cycle.¹⁷

The most widely discussed example of such an autopoietic system is a certain bacterium which, as a result of chemotaxis, moves towards an increasing level of concentration of a chemical substance (e.g., glucose). This organism serves as an example of the normativity of autonomous systems that act to achieve a state that is “good” for them. So the bacterium is a minimal autonomous mechanism, but is this enough to make it count as an agent – i.e., the author of an action? How is it different from a hurricane or a candle flame, which also form, in each case, a dynamic unity that interacts with the environment but which we would definitely not consider agents? To answer these questions, it is necessary to specify the conditions for autonomous agency. This is also essential if we are to determine whether extended cognitive systems and groups may be autonomous agents.

3. From Autopoietic to Sensorimotor Autonomy

As I have already mentioned, in analytical philosophy of action, autonomy is seen as necessarily linked to the capacity for reasoning and reflection, where this is vested exclusively in humans. Yet functionalist cognitive science seeks to explicate sophisticated cognitive abilities as properties belonging also to cognitive systems that are not humans as such. Nevertheless, philosophically minded researchers are focused not so much on the actions that an autonomous system is capable of undertaking, as on the processes that serve to realize the property of autonomy – in other words, the conditions of its possibility.¹⁸ This is precisely the approach pursued by Maturana and Varela, and continued and refined by other researchers. One of the conditions through which hurricanes and candle flames can be distinguished from the bacterium mentioned in the preceding

¹⁴ Maturana, Varela (1980).

¹⁵ Varela (1991).

¹⁶ Varela (1979).

¹⁷ Maturana, Varela (1992).

¹⁸ Moreno et al. (2008).

paragraph was provided by Howard H. Pattee. In his view, compensating for the energy losses that are necessary for an autonomous system to remain in existence is only possible if the system has the ability to select between alternatives. Action only occurs where there is a choice; otherwise, motion arises from successively unfolding states of some law-determined temporal sequence of events. In other words, an autonomous system not only reacts to external events and processes, but is also capable of behavior that originates from itself. Such a thesis, however, presupposes the existence of a causality beyond the dynamics of what is law-governed: namely, one whose source is the system itself. This interesting solution – which, however, goes too far beyond the scope of this paper to be presented here – is analyzed in detail by Pattee in his article “The physics of symbols: Bridging the epistemic cut.”¹⁹

According to those who consider autopoiesis essential for autonomous action, the agent furnishes the source of its own unity (identity) and maintains the latter by introducing appropriate changes into its environment. An autonomous system is therefore a kind of autopoietic machine, the existence of which depends both upon the internal processes that constitute its identity and upon its interaction with the external world. The more complex both of these types of process are, the more autonomous the system becomes. It has, in a word, greater control over adverse changes in its environment, and so is more independent of them. In this sense, autonomy is sometimes considered gradable.²⁰ It should be noted however, that that is not the generally accepted position on autonomy among enactivists.²¹ Ezequiel Di Paolo, among others, considers autopoiesis as an all-or-nothing property that does not come in degrees, and he introduces the concept of adaptivity as gradable and essential for autonomous cognitive agency.²² This is an important voice in the discussion on autonomous agency because it clearly indicates the difference between the original autopoietic understanding of autonomy introduced by Varela and Maturana, and the concept of sensorimotor autonomy developed later by enactivists. It is worth taking a closer look at this distinction, as it will allow us to better understand the conditions that a non-living system must meet in order to be an autonomous agent.

Di Paolo indicates the difference between a barely autopoietic system that is only self-productive and self-distinctive and a system that is adaptive and capable of sense-making operations. The very identity of the system is given through autopoiesis, yet there could be a case of a non-adaptive conservation of autopoiesis within the system, the self-production of which is unaffected and which does not have any compensatory mechanisms that maintain its identity.²³ Namely, adaptivity and normativity are not included in the definition of an autopoietic system, hence theoretically, there can be a system that is autopoietic yet non-adaptive. According to Di Paolo, only a system that could improve its situation and maintain its existence in the face of unfavorable conditions is an autonomous cognitive agent. Such a system is able to regulate its behavior according

¹⁹ Pattee (2001).

²⁰ Moreno et al. (2008).

²¹ I am grateful to an anonymous reviewer for bringing this to my attention.

²² Di Paolo (2009); Beer, Di Paolo (2023).

²³ Di Paolo (2009).

to the norms that it itself generates; in other words, it is adaptive. Being adaptive, such a system is sense-making. It regulates its coupling with the environment, under precarious circumstances,²⁴ by differentiating the solutions that are better for maintaining its autonomous identity. A system that is autopoietic and adaptive in this way can be defined as an autonomous agent.

Di Paolo's approach to the relation between autopoiesis and autonomy is shared by other enactivists. Xabier Barandiaran, among others, argues that the concept of autonomy as a material autopoiesis ignores other levels of autonomy, for instance, sensorimotor autonomous systems.²⁵ Enactivists explain the autonomous cognitive agency of such systems with reference not to their autopoietic nature (which they most often assume as a necessary and obvious condition), but rather to the their sensorimotor dynamics, namely to the emergence of its adaptive behavior out of the interaction it enters into with its environment. Hence, an autonomous agent not only has to be self-distinctive, but it also has to be the source of its own activity and normativity, so that no external entity imposes on it rules of behavior nor controls its compliance with them.²⁶ Adaptivity and sense-making are also the defining features of social autonomy. In cases of group autonomous systems, their participants enter into processes of coordination and mutual perturbation that result in joint meaning-making and group autonomous agency.²⁷ I refer to this argument in the Section 5, where I undertake the issue of group autonomous systems.

4. The Enactive *Mark of Agency*

Investigations of the mechanism underlying autonomy, as largely pioneered by Maturana and Varela, aim to respond to the need for ontological solutions that will shed light on the intrinsic characteristics of autonomous systems. A system that merely behaves like an autonomous one, but does not have a corresponding internal mechanism defined by coupled causal processes, cannot be considered an agent on the basis of this approach. So what are the conditions that a mechanism responsible for genuine autonomy must meet? A detailed and convincing proposal for this sort of distinguishing *mark of agency* has been presented by Xabier Barandiaran, Ezequiel Di Paolo, and Marieke Rohde.²⁸ These researchers identify three conditions that a system must meet in order to be considered an autonomous agent: firstly, it should constitute a distinct individual; secondly, it has to be the source of its own activity relative to the environment; and thirdly, it needs to regulate its activity according to certain norms. The authors in question call these con-

²⁴ The condition of precariousness is defined as follows: "in the absence of the enabling relations established by the operationally closed network, a process belonging to the network will stop or run down [...] a precarious, operationally closed system is inherently restless, and in order to sustain itself despite its intrinsic tendencies towards internal imbalance, it requires energy, matter, and relations with the outside world. [...] Hence, the system is not only self-enabling, but also shows spontaneity in its interactions due to a constitutive need to constantly "buy time" against the negative tendencies of its own parts." Di Paolo, Thomson (2014): 72.

²⁵ Barandiaran (2017).

²⁶ Thompson (2007).

²⁷ De Jaegher, Di Paolo (2007).

²⁸ Barandiaran et al. (2009).

ditions *individuality*, *asymmetry*, and *normativity*. In a word, it must be that the agent, as an individual, *does something by itself* according to certain norms in a given environment. The presented definition of agency is in line with an enactive, sensorimotor account of autonomy, which considers autopoiesis as a necessary but insufficient condition.

Each of the above-mentioned conditions raises problems and requires clarification. The first one is related to the question of a system's boundaries, as these can be set by an observer quite arbitrarily, depending on current needs. The genuine boundaries of the system should therefore somehow be distinguished from those identified by an observer interpreting its behavior from a specific perspective (e.g., from the intentional stance). Yet in the case of artificial systems, there are no such authentic limits, as their individuation is always implemented by their designer.²⁹ If a system cannot independently define itself as a separate individual, then it cannot claim to be an agent – at least where the conditions stipulated by Barandiaran et al. are concerned. The crucial question, then, is whether only a living system is capable of self-organization and self-identification – and if the answer is positive, why that is so. An attempt to answer this question leads to the second and third conditions mentioned above. According to the second one, which the authors call *interactional asymmetry*, an autonomous system cannot merely react passively to external forces, but must itself be the source of its activity. The system, so to speak, enters into asymmetrical coupling with the environment, in that it introduces changes that work to its own advantage. To do so, it must acquire and accumulate the energy necessary for behavior that will sustain its existence. The agent performs all of these activities because existence is valuable to them – an observation that then directs thinking about agency towards the condition called *normativity*.

The close connection between these two defining features of an autonomous agent – interactional asymmetry and normativity – is explained by Barandiaran et al. using the example of a person suffering from Parkinson's disease. Such patients are not considered to be the agents of their spasms, despite the fact that they are individual systems, and the source of their tremors is within themselves. By contrast, a genuine action should realize some intrinsic value entertained by the agent, leading to its self-maintenance. Tremors are caused not by the pursuit of self-preservation as a value, but by the dysfunctionality of the body's subsystem. If the system's behavior is aimed at self-preservation, it can be judged as successful or not. The most important point here, however, is that this normative aspect of an autonomous action could not be assigned from the outside by an observer, for whom some particular behavior on the part of the system may or may not be seen as beneficial. It must be the internal organization of the system itself that determines the conditions for its survival. In short, an autonomous system should undertake actions for itself, and not for either an observer or a designer.³⁰

All three conditions of agency cited above characterize the internal organization of an autonomous agent and how it interacts with its environment. Contrary to the intentional stance, this strategy allows the system to be assigned agency not as a result of merely observing its behavior, but rather by examining its internal properties. By satisfying the three conditions, a system becomes an autonomous agent without the need

²⁹ Jonas (1968).

³⁰ Barandiaran et al. (2009).

for conceptual language, propositional attitudes, intentions, self-awareness – in other words, without a mind. It is enough to be an autonomous organization of coupled interdependent processes capable of adaptively regulating interactions with the environment according to norms determined by its own survival-related conditions.

At this point, it is worth noting the danger of a vicious circle that, in principle, threatens all definitions of agency: an agent is defined by terms such as “doing something,” “taking up activity,” and “acting,” but these behaviors already assume the concept of an agent, because there must be an entity that performs them. Yet what does it mean to act, to do something? Barandiaran et al. assert that their definition of agency is free of this circularity, as the second condition states that “an agent is a source of activity, not a passive sufferer of the effects of external forces.”³¹ Moreover, the authors in question specify this condition by pointing to a system capable of modulating its coupling with the environment in an adaptive manner: “where *modulation* indicates an alteration (...) in the set of constraints that determine the coupling between S (the system) and E (the environment).”³² Hence, being a source of its activity and behaving in an adaptive manner distinguishes agents from non-agents.

Davidson and other representatives of the analytical philosophy of action point to intentions (mental states) as making a difference in this respect. In the propositions just discussed, however, the difference seems to lie in the fact that the source of the movements we call action lies within the subject, so that such movements are not caused by forces acting directly from outside. Moreover, as the example of the person suffering from Parkinson’s disease shows, these movements must fulfill an adaptive function: they can, admittedly, fail in this regard, but in principle they must be directed at achieving some goal that contributes to the system’s existence and development. Even so, distinguishing inaccurate movements that are nevertheless adaptive in principle from intrinsically dysfunctional ones proves quite a challenge – if it is possible at all. The question of what this inner source of action is, and what the ability to choose actions on the part of even the simplest of organisms may consist in, presents us with an extremely interesting issue as regards different types of causality. That, however, is a subject for another paper.

In concluding the present analysis of the concept of an autonomous agent, it is worth noting the ambiguity inherent in the attempt to divide research on agency into those adopting the intentional stance approach and those who seek genuine agency through an examination of the properties of the system’s underlying mechanisms. Even if we embrace the self-individuation requirement which says that the system must define its own boundaries itself in order to function efficiently with a view to preserving its unity (identity), those investigating this system must still somehow recognize these boundaries from their perspective. But how can the mechanism that constitutes the identity of the system be determined from the observer’s perspective? Is it always clear and intuitive? Will the agent always simply be an individual organism? Enactivists emphasize the role of the observer in identifying autonomous agents, arguing that systems that meet the conditions of sensorimotor agency (presented above in reference to Barandiaran et al.) can

³¹ Barandiaran et al. (2009): 370.

³² Barandiaran et al. (2009): 376.

be distinguished in various ways, depending on the perspective set by one's specific research goals. This position combines, in a way, the intentional stance approach with that of recognizing autonomous agency by examining features of the system's underlying mechanism. Autonomy is, namely, co-constructed by the system's mechanism, the observers that identify it, and the system's environment; it is a quality that emerges in the mutual interaction between all of these elements.³³ Hence, the observer of a given system should utilize both strategies. On the one hand, being reliable, they can evaluate the system's agency by examining its functional organization, but on the other, they can also define the framework of the mechanism under examination quite freely, by including or excluding specific processes depending on the perspective chosen.

This is precisely the problem recognized by those enactivists who point to the difference between purely autopoietic systems (minimally autonomous through organizational closure) and sensorimotor autonomous systems. They recognize autonomous agency where the adaptability of the behavior is evident: in other words, where the behavior of the system is directed towards the value of maintaining unity, and where this has not been conferred on the system by another agent, but stems from the sheer fact of its being alive. Fred Cummins aptly comments on this enactive position when he remarks that "with life, value leaks in."³⁴ Hence, an autopoietic system should not be confused with autonomous, sensorimotor agency. It need not be an adaptive and sense-making agent, so the observer can set the limits to an autopoietic mechanism quite freely, subject to the research perspective adopted. By contrast, the attribution of autonomous agency goes further. In order to ascribe it to a given system, it is necessary to recognize the value that is rooted in the identity of the system itself – in its conditions of existence, so to speak. This, however, according to enactivists, poses a serious problem, as depending on the perspective adopted by the observer, different values may be perceived, and may be implemented by differently defined mechanisms. Cummins illustrates the dependence of the boundaries of an agent on the observer's perspective with the example of a newborn kitten whose mother moves it closer to her own body to keep it warm. This scenario was cited in the work of Barandiaran et al. as an example of a system that does not constitute an agent, in that the source of the kitten's movement is not itself but its mother, which constitutes a separate system within the environment.³⁵ If, however, as argued by Cummins, the observer changes perspective and treats the feline family as one system, the source of its movement will be internal, and all three conditions of agency will be satisfied, so there would seem to be no reason why such an extended system should not be considered an autonomous agent.³⁶ This does not mean that Barandiaran et al. are mistaken in ascribing agency or defining the system; it only means that the mechanism responsible for the constitution of the system may be differently determined depending upon the perspective chosen by the observer. Adopting this thesis opens up a wide range of new ways of explaining the autonomous agency of systems recognized by researchers at different levels of reality.

³³ Sanches de Oliveira et al. (2023).

³⁴ Cummins (2014): 109.

³⁵ Barandiaran et al. (2009).

³⁶ Cummins (2014).

With the difference between an autopoietic system (organizationally closed) and an autonomous (adaptive, sense-making) agent defined, the question of whether extended, group, and artificial cognitive systems can be autonomous is posed more precisely. After a brief reflection on cognitive artifacts, as an in-depth analysis of their autonomous agency is beyond the scope of this work, I will focus on the autonomy of extended and group systems.

First of all, to be autonomous in the enactive sense, cognitive artifacts would have to be systems that are not fully, directly designed, but to which a significant space for self-production and self-maintenance has been left. In order to model them, computer scientists create algorithms that, based on limited data, emerge at some stage in their development as complex processes, simulating the natural phenomena of learning and evolution. As a result, the dynamics of such algorithms become much more complex than the basic rules introduced by the programmer, and cannot be reduced to them. Models of neural networks, immune systems, or robot communities that learn are examples of such programs. All of them mirror the development of a natural autonomous system which, from simple initial conditions, constitutes a complex unity that develops and learns in response to changing environmental conditions. Yet do such algorithms exhibit genuine autonomy, comparable to that of natural systems? If not, what is the difference? Clarification of these issues cannot be achieved without extensive work on computational simulations and robotic models, as well as research in biochemistry and biotechnology. Constructing artificial agents promises not only to contribute positively to the development of the latest technology, but also to enable a better understanding of the mechanism that realizes the property of autonomy and its relation to life and agency. Below I present analyses concerning the autonomous agency of systems that are composed of humans, yet they are not individual humans as such, and so they are not alive as a single entity. I believe that, apart from being interesting in itself, this discussion provides a valuable basis for considerations regarding the autonomous agency of artificial systems.

5. Extended Cognitive Systems as Autonomous Agents

When evaluating human movements as either actions or non-actions, it is important to clearly distinguish the level of the organism from that of the person. In Davidson's view, actions appear only on the personal level, as they must be preceded by an intention. Thus, on the basis of his philosophy of action, a physical organism as such, or an artificial system, cannot be an agent, as this status is granted exclusively to human beings with minds. Enactivists, however, using a broader definition of an agent, allow for the possibility that the subject of an action could be an organism in the sense of some physical mechanism realizing systemic mental properties. Which movements performed by the human body constitute an action depends on which system is being considered. The unconscious processes in a person's body are the actions of an agent – namely, a biological organism that realizes the goal (and value) of survival. All the same, many human activities are difficult to explain normatively from a purely somatic perspective. Writing a book, watching a movie, or voting in an election do not directly accomplish

the goals that can be attributed to the mechanism constituted by the coupling of adaptive physical processes. It seems more accurate to attribute this to the person, as a higher-level system constituted by a coupling together of mental processes directed toward a variety of human goals. Moreover, there are situations where the individual human being is only one part of a wider cognitive system.

The idea of extended cognitive systems was introduced by Andy Clark and David Chalmers in their famous article “The Extended Mind”³⁷ and has since been further developed in many other works by Clark.³⁸ Clark and Chalmers analyze cases where a person uses an artifact in order to solve a cognitive task. The extended mind thesis states that if there is a relation of continuous reciprocal causation between the processes taking place in the human body and those occurring in the artifact, then these will together constitute a cognitive process of sorts, and so can be viewed as a single cognitive system. Cognitive artifacts, which are external to the human body, affect and often change our natural (human) cognitive abilities. A person, in turn, manipulates the artifact (in terms of its inner workings), and in this way such causal reciprocity comes to be realized. The most common instances of such artifacts are smartphones, with their diverse applications, and computers running software programs. When using these, the agent offloads a certain amount of cognitive work onto the artifact, and it determines his or her future actions in turn. Each step in this extended cognitive process depends on the preceding ones, and it is almost impossible to distinguish the internal and external parts of it. Such a process is realized by a single internally coupled cognitive system. But can this be recognized as an autonomous agent in the sense presented in the foregoing paragraphs?

I do not see any serious objections to regarding a one-person extended system as an autonomous agent. However, an assessment ought to be made of whether it meets the adopted conditions of agency. Let us assume that these are the conditions presented by Barandiaran et al., supplemented by Cummins’ thoughts regarding the influence of the observer on the boundaries of the system under consideration. Thus, as suggested by enactivists, I propose combining both of the strategies for studying agency as it pertains to systems: i.e., the approaches that appeal to intentional and to “genuine” (intrinsic) considerations, respectively. Before continuing this line of thought however, it is important to notice the tension in the combination of the functionalist approach with enactivism, which takes a strongly embodied position towards cognition.³⁹ Both Dennett’s intentional strategy and Clark and Chalmers’ extended mind thesis are rooted in functionalist view on cognition as high-level problem solving, which neglects the physical details of the body that realizes this process. Although the system’s body is, according to Clark, cognitively significant, he defines it in reference to certain functional roles,⁴⁰ and whatever plays these roles should be recognized as the body, regardless of its physical or biolog-

³⁷ Clark, Chalmers (1998).

³⁸ Clark (2008, 2010).

³⁹ I am grateful to an anonymous reviewer for bringing this problem to my attention.

⁴⁰ These roles are defined as that of being “the locus of willed action, the point of sensorimotor confluence, the gateway to intelligent offloading [for problem-solving computations], and the stable (though not permanently fixed) platform whose features and relations can be relied upon in the computation of certain information-processing solutions.” Clark (2008), 55–56.

ical properties.⁴¹ Such a functionalist conception of cognition is seriously contested by enactivists, who recognize the physical properties of the system's body as fundamental for its cognitive abilities, since cognition emerges out of the interaction between the body and its environment. Recall that enactivists define cognitive autonomous agency as operational closure of a system's constituent processes that generate and maintain its identity and adaptivity.⁴² To be an autonomous agent, a system has to enter into normative engagement with its environment. Only from such an interaction could original, not derivative, meaning emerge. Cognition, according to enactivists, always occurs in such a relational domain, namely it is the relational process of sense-making that takes place between the system and its environment. As such, cognition does not have a location, hence there is no point in considering whether it is realized inside the brain or partly outside of it.⁴³ Enactivists simply do not take part in the internalist/externalist debate. Nevertheless, they do engage in the functionalism/embodiment one, and they strongly favor the latter. Having said this, employing the intentional strategy and evaluating the system's agency from the observer's perspective may seem not quite correct, as far as the enactive conception of autonomy is concerned. I do not recognize it as an inconsistency, however. Systems that meet the enactive conditions of agency can be distinguished in various ways, depending on the perspective set by the observer's specific research goals. Claiming this, enactivists do not pay secret tribute to functionalism but rather notice the diversity of autonomous systems around us, some of which may overlap. It is the observer's concern to differentiate the system that he or she evaluates in terms of autonomy, and that can be done in various ways. Let me now return to the question of whether a one-person extended system can be an autonomous agent as far as the enactive conditions presented by Barandiaran et al. are concerned.

The extended one-person cognitive system that meets the condition of continuous reciprocal causation between the human and the artifact manifests self-organization as a separate object, in the sense that it sets its own boundaries. Admittedly, this condition is only realized by one part of the system, the human being, but the question of what specific processes taking place in the extended system serve to determine its autopoietic autonomy is not important for its evaluation in this respect. When a person enters into a reciprocal causal relation with an artifact, they and it together constitute a new object: an extended cognitive system. Moreover, they do so in order to achieve something possessing a certain value that they could not have accomplished on their own. The motivation to act, intention, and responsibility for acting belong to the human being, who is the only reflective and conscious part of the extended system. However, when applying the conditions for agency grounded in autopoiesis and adaptivity proposed by enactivists, the agent will be the entire extended system, as enactivists invoke either intention as a mental state or the notion of responsibility. The system in question will count as a separate entity in virtue of whatever mechanisms are involved in working to perform a specific task, where this includes processes taking place in both the human body and the artifact.

⁴¹ Clark (2008).

⁴² Thompson, Stapleton (2009).

⁴³ Di Paolo (2009).

Similarly, it is possible to justify the claim that such an extended system can meet the conditions of asymmetry and normativity: it is the source of its own activity in the environment and regulates its actions in order to achieve something of value, the source of which is also to be found in itself. Once again, it is the human being that is responsible for fulfilling both of these conditions, in that he or she possesses an intention that causes action. Yet the processes that realize this are distributed over the entire extended system. The exact location of the elements of initiative and control is irrelevant to this conception of agency. The coupling between human and artifact, however, can be considered from more than one perspective. The actual person who, for example, uses a sensory substitution device or a computer program to perform specific cognitive tasks most often treats himself or herself as a separate entity, using a different system for his or her own purposes. However, viewed from the perspective of the task itself, one can readily see that the processes constituting the extended mechanism responsible for cognitive success are so closely coupled with each other that it makes sense to regard them as making up one single whole – human-plus-artifact – created in order to achieve something specific in respect of value. Such an extended system constitutes its own boundaries and is also the source of its own activity and normativity; for this reason, and relative to the conditions for agency discussed here, it will count as an autonomous agent.

On the definition of agency put forward by Barandiaran et al., a system qua agent exhibits the essential features of a living autopoietic system to which Varela devotes his analyses. But does it have to be alive to count as autonomous? The extended cognitive system that could be considered an agent is not alive as such: only the human part of it is living. Thus, it seems that life is not a necessary condition for minimal autonomy (defined through the notion of organizational closure). The conditions presented by Barandiaran et al. do not themselves make an agent count as alive. On the other hand, it is the living person who is actually responsible for the extended system's fulfillment of the conditions we are discussing. Without him or her, all acting with reference to some intrinsic and genuine locus of value seems impossible – at least from the standpoint of today's level of technological development. Perhaps, however, this is a case of biocentric thinking, in that many computer scientists are inclined to call evolutionary algorithms, or those that model the immune system, "living," just because they exhibit certain characteristics essential for life, such as the ability to replicate, mutate, exchange information, evolve, and adapt to their environment.⁴⁴ The thesis that biological life is not necessary for autonomous agency is also supported by analyses of groups as agents. Below, I consider a number of the most convincing arguments in favor of group agency.

6. Group Systems as Autonomous Agents

Can a group be an autonomous agent in an enactive sense? In order to be such, according to the argument presented in this paper, it has to meet the conditions of autonomous agency presented by Barandiaran et al., and to do that it has to be self-productive and self-sustaining. Philosophers engaged in studying this issue find themselves in dispute

⁴⁴ Adamatzky, Komosinski (2009).

with each other when it comes to considering whether the concept of autopoiesis, introduced by Varela and Maturana, allows for such an interpretation. It is worth reminding ourselves here that the autopoietic system is defined by the notion of operational closure, which means that it is organized in such a way that it produces its own components and maintains its own existence. As Maturana explains:

a composite unity whose organization can be described as a closed network of productions of components that through their interactions constitute the network of productions that produce them, and specify its extension by constituting its boundaries in their domain of existence, is an autopoietic system.⁴⁵

The autonomy of an autopoietic system lies in the fact that its operation is determined solely by its own internal states and organizational structure. External events can affect it only to the extent of disrupting its functioning, which stimulates the system to make changes in this structure and its internal states – changes whose effects are not then controlled by any external factor.

Maturana and Varela's theory is directed at biological systems in which cognitive processes are carried out by a neural system exhibiting organizational closure. Neural activity, in short, is both triggered by and results in other neural activity, in a closed cycle.⁴⁶ According to the authors, this is one of the features that distinguish living systems from all others, so it would seem that groups, qua non-living entities, cannot be autopoietic. On the other hand, one could construe autopoietic organisms as just one type of autonomous system and try to show that non-biological systems also satisfy the condition of operational closure. Such a strategy is adopted by proponents of the idea that we should treat certain groups as autonomous agents, i.e., as systems which are self-productive and self-maintaining. But is it at all possible for such a circular, closed, dynamic process of self-creation to occur in a group?

The crucial difference between biological and group systems concerns their components, which in the second case are reflexive agents. In contrast to the components of biological systems, the actions of conscious group members do not depend solely on physical interactions but can be guided by non-physical, linguistic factors, such as their knowledge of the behavior of the group as a whole. Members' knowledge of the overall state of the group can, in other words, have a direct impact on the behavior of its parts, which does not happen in biological systems. Hence, group systems exist through the interaction of two domains – the physical and the non-physical.⁴⁷ While humans, as organisms, exist in the physical space, groups exist through the interaction of physical and non-physical (linguistic) domains. They are products of self-organization among

⁴⁵ Maturana (1987): 349.

⁴⁶ Maturana, Varela (1992).

⁴⁷ By "non-physical domain" I do not mean Cartesian mental substance, but an emergent, systemic property of the system that is not reducible to the physical properties of its parts and the laws governing them. Language is understood here as a property that has emerged out of both the biological organizational complexity of human organisms and the complexity of cultural and social interactions between them.

their constituent autopoietic unities. Accordingly, people, being members of a group, are themselves complex autopoietic systems that manifest, via their use of language, a capacity for reflective self-awareness and control with respect to the operation of the larger system to which they belong. They can consciously choose to enter into a relationship with other people that make up the structure of a group system.⁴⁸ Such a system can be considered organizationally closed, as the unity and the persistence of this whole are determined by the internal relationships between its members, created, maintained, and controlled by the group itself. The processes of self-production and self-individuation characteristic of an autopoietic system are partly reliant, in the case of a group, on linguistic interaction between its components, which is unique among other systems of this type, where the mechanism of autopoiesis is realized only by physical processes.

In addition to organizational closure, the autonomous system manifests its own internal normativity, which means that the goals guiding its actions are not imposed by another agent but are inscribed in the internal structure developed in the process of self-creation. The norms that guide the actions of a group constitute its organizational structure, which in turn determines the relations between group members. To repeat, these norms are not given to the group from outside, but are formed in the dynamic process of self-production of the mechanism that constitutes the group as a separate whole, in which linguistic communication plays an important role. Humans and other autopoietic systems maintain their own existence via appropriate physical mechanisms. In contrast, groups depend not only on physical interactions (keeping their components in existence) but also on non-physical linguistic interactions.

Chris Goldspink and Robert Kay – researchers participating in the discussion concerning the autonomy of group systems – argue that an effective strategy for analyzing group agency is to combine the concept of autopoiesis with dynamical systems theory.⁴⁹ The concept of autopoiesis makes it possible to explain the rationality and normativity of the group system without having to refer to an external conscious rational agent. Combined with the tools of dynamical systems theory, Maturana and Varela's analysis provide a picture of autopoietic individual agents that, by entering into complex interactions with each other, produce an organization that constitutes a single emergent system manifesting rationality, goal-directedness and, in consequence, autonomy.

Within the debate on group agency, there are also critical voices seeking to highlight the ontological problems raised by the thesis that a group system can be organizationally closed. John Mingers argues that an autopoietic organization is composed of elements that are themselves produced by that system in the process of self-creation and form the boundaries of the system, distinguishing it from elements that already belong to its environment. Human beings, on the other hand, as group members, are not produced by that system itself, but are the product of other physical and biological processes that cannot be directly included in the mechanism responsible for keeping the group system in existence. Besides, it is difficult to view people as the elements constituting the unity of a group and responsible for defining its boundaries, as they can freely decide to belong

⁴⁸ Goldspink, Kay (2003).

⁴⁹ Goldspink, Kay (2003).

to or leave a given group – and, what is more, can belong to many different groups at the same time.⁵⁰ So what, then, constitutes the boundaries of the group system, and on what basis can it have the ability to act attributed to it, given that the physical agents involved are only individual persons?

The solution to this difficulty may well lie in a different definition of the components of the group system. Such a strategy is employed by Niklas Luhmann, whose contribution is considered a classic in the context of the debate under consideration here.⁵¹ While he recognizes various types of social systems as autonomous agents, he does not treat their individual members as responsible for the system's autonomy. Luhmann argues that group systems are organizationally closed and self-producing, but what creates and determines their unity and existence are not physical elements and processes, but linguistic communication. By "communication," he means an event consisting of three elements: information, utterance, and understanding. Thus, it occurs only when the recipient receives and interprets the information transmitted.⁵² Moreover, the system itself determines what information falls within its framework, and how it can be expressed and interpreted. Outside of such systems, information does not exist, because the physical environment as such does not communicate with anything. The physical events occurring in it can only affect the group system when they become an object of communication, so it cannot communicate with the environment, but only about it, and within its own information processing capabilities. To sum up, Luhmann's theory is an attempt to apply the concept of autopoiesis to a non-physical system, the basic components of which are not physical objects but instances of communication.

Mingers, however, has doubts about the interpretation of group systems proposed by Luhmann. He sees a problem with the phenomenon of *production*, which in Luhmann's approach is realized not by people, but by communication events that engender subsequently such events, constituting a closed cycle responsible for the process of autopoiesis. According to Mingers, Luhmann does not sufficiently explain where these come from, or how they arise from interactions between individual members of the group.⁵³ While it is hard to deny that without human activity there would be no communication, Luhmann does not show us the relationship between mental, conscious, individual systems and the group communication system. In short, he focuses exclusively on the role of communicative events *in abstracto*, passing over the problem of real individual-society relations in his theoretical reflections.

The analyses undertaken by Luhmann lead to the conclusion that the process of self-production and self-maintenance need not take place only in living systems, as other systems are also able to create forms of genuine normativity that determine the purpose of their activity. Apart from consciousness, the basis for creating something of value in an autonomous system ranges over communicative events themselves – thus, for Luhman, self-production, self-maintenance, and intrinsic normativity should be

⁵⁰ Mingers (2002).

⁵¹ Luhmann (1986).

⁵² Luhmann (1995): 137.

⁵³ Mingers (2002).

thought of as independent of being alive.⁵⁴ Nevertheless, Mingers' remark pointing to the importance of the participation of conscious individuals in the creation of communication events raises doubts about the possibility of an autonomous system that does not manifest life. Although the group, as such, is not a living system, it is composed of living entities that enable the existence of the components of the group system responsible, according to Luhmann, for its autonomy. Yet acknowledging this point does not itself solve the problems related to self-production and the boundaries of the group system highlighted by Mingers.

Even though groups are not living systems as such, this is not tantamount to excluding them from the class of autonomous agents. Recall that the definition of an agent proposed by Barandiaran et al. makes it possible, at least theoretically, to register systems that do not manifest the properties of life and consciousness as agents. Most importantly, any such system should be autonomous, where autonomy is understood as interactional asymmetry, corresponding to its being the source of its own activity in the environment. This means that the system must acquire on its own the energy necessary for its survival, where this is the driving force that enables it to change the environment to its advantage.

So can a group system have, as such, its own driving force? Can it, in other words, as a separate entity, be causally effective? Philosopher Dave Elder-Vass argues that it can, pointing to the emergent nature of group agency.⁵⁵ In this instance, "emergent" means having a causal power different from that of its underlying basic properties. Only with this assumption in place is it possible to maintain the thesis of the autonomy of group systems. Moreover, the harsh criticism directed against the idea of group agency by reductionists does not seem to discourage Elder-Vass. In his opinion, even assuming the possibility of explaining emergent properties in terms of more basic ones, it is possible to grant certain groups their own causal powers. The relationships between group members constitute the mechanism from which the group emerges as a new whole with new properties, including new causal powers. A reductive explanation of this process does not make these powers disappear. It is hard not to agree with him that without such an assumption it would be difficult to attribute causal forces to anything in the world, for the action of almost any object can be explained in terms of the causal interactions between its elements.

Elder-Vass also comments on the normativity of group systems, which is, according to the definition he adopts in his work, another necessary feature of agency. As has already been said, the system that acts as an agent is meaning-productive in the sense that the values it pursues by acting in the world come from within it and not from outside it. As an example of the normativity of a group system, Elder-Vass points to the so-called *norm circle* formed by people jointly involved in disseminating a given norm of behavior. Members of the norm circle enter into relationships that generate a collective power to

⁵⁴ It is worth noting that for enactivists life itself is an emergent property of a network of physical processes organized in a particular manner. Yet, what makes groups autonomous systems is not a network of physical interactions between their members, but a network of linguistic interactions between them, which are not by themselves enough for the emergence of life.

⁵⁵ Elder-Vass (2010, 2014).

influence human behavior both inside and outside of it – a power which none of the group members possesses individually. The organizational structure of a given group is grounded in the norms that determine its functioning, which the group establishes for itself and whose implementation it controls. Elder-Vass demonstrates the emergent causal power of social conventions using the example of the queue: the latter forces individuals to behave in a specific way – they must comply with the norms in place to achieve their desired goal. The queue's participants, and the relations obtaining between them, produce an emergent property that has the power to induce specific behavior. This property is not possessed by any participant of the queue individually, nor would it be by a group organized in any other way.⁵⁶ The queue is therefore an example of a social phenomenon with an emergent casual power grounded in specific norms that steers the behavior of individual people. Importantly, the beliefs of the queue's members about how they should behave are determined by a norm that is not imposed externally but rather produced by the queue itself.

Although the queue is a spontaneously formed group, and it is difficult to ascribe to it a goal whose implementation could furnish a reason for its existence, this phenomenon is an example of top-down normative causation – the basis for granting some groups the status of an agent. Any group formed intentionally by individual people is created with a view to the furtherance of a particular value or set of values. The norms determining the behavior of its members constitute the organizational structure that is the mechanism of its functioning. This mechanism, by acting, renders all the decisions, beliefs, and aspirations of a given group, including the norms that create it, for they do not come from outside, but rather are its effect while at the same time being its cause. This is the self-creation of an autonomous, normative system, which thus acquires the status of an agent.

Analyses concerning the autonomous agency of group systems lie also at the core of enactive theories of social cognition. Such theories are concerned with defining the social interaction between individuals in terms of joint sense-making and its experience. To repeat, group systems are distinguished from other autonomous systems by the fact that patterns of coordination directly influence the actions of individual group members who are involved in sustaining and modifying group identity and who themselves remain autonomous. Social interaction between group members generate and maintain an identity of a group system that exhibits precariousness, operational closure, and participatory sense-making which are, according to enactivists, defining properties of a social autonomous agent.⁵⁷

7. Conclusion

Barandiaran et al., as well as Varela and Maturana, tend to recognize only living systems as autonomous agents. They argue that currently only living organisms, and every one of them from the simplest to the most complex, are self-sustaining, which means that they

⁵⁶ Elder-Vass (2010).

⁵⁷ De Jaegher, Di Paolo (2007); Di Paolo, Thompson (2014).

are able to self-maintain their existence as a result of drawing energy from the environment using their own resources. In writing this article, I am motivated by the question of the autonomous agency of cognitive systems that, as a whole, are not human beings. I have compared the standard and enactive conceptions of agency, of which only the second can be applied to systems that are not humans as such. The first position, found in the analytical philosophy of action, is built around the concept of intention, while the second refers to the theory of autopoiesis and sensorimotor agency. All living organisms are autopoietic, and therefore can be ascribed some minimal agency, but can this quality be attributed to systems that do not exhibit the property of being alive? In order to answer this question, I invoked the conception of agency presented by Barandiaran, Di Paolo, and Rohde, based on three conditions that assume Maturana and Varela's definition of autopoiesis and yet are also capable of being met by a non-living system. I applied these conditions to the analysis of an extended system composed of a person and a cognitive artifact, and I concluded that it could indeed satisfy them. Then, I adduced arguments in support of the autonomous agency of certain group systems, which are also not alive as such. Despite the many uncertainties in this regard, my conclusion is also positive in this case. Both kinds of cognitive system under consideration, however, are composed of living parts that are directly responsible for fulfilling the conditions of normativity and interactive asymmetry which, together with individuality, constitute the core of autonomous agency. Nevertheless, biological life is not necessary for a cognitive system to be autonomous, for autonomous agency could emerge out of processes other than those responsible for life. Linguistic processes taking place in group systems are examples of these. Niklas Luhmann points to linguistic communication as non-physical interaction between humans, out of which an autonomous system may emerge. Enactivists, meanwhile, define social interaction in terms of participatory sense-making emerging in intentional activity in interaction, which also could be interpreted as non-biological.⁵⁸ Hence, I do not wish to rule out the possibility that one day an autonomous artificial agent will be created, as technological developments have on more than one occasion exceeded the expectations of even the boldest imaginations. Such a system may not be alive in a biological sense, yet it is likely to manifest the functional characteristics of a living organism, and for this reason could well be characterized as an instance of artificial life.

Funding: None.

Conflict of Interest: The author declares no conflict of interest.

License: This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

⁵⁸ De Jaegher, Di Paolo (2007).

References

- Adamatzky A., Komosinski M. (2009), *Artificial Life Models in Software*, Springer, London.
- Barandiaran X. E. (2017), "Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency," *Topoi* 36: 409-430.
- Barandiaran X. E., Di Paolo E., Rohde M. (2009), "Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action," *Adaptive Behavior* 17 (5): 367-386.
- Beer R. D., Di Paolo E. A. (2023), "The Theoretical Foundations of Enaction: Precariousness," *Biosystems* 223, 104823.
- Chisholm R. (1976), "The Agent as Cause," [in:] **Action Theory**, M. Brand, D. Walton (eds.), D. Reidel Publishing Co. Dordrecht: 199-211.
- Chopra S., White L.F. (2011), *A Legal Theory for Autonomous Artificial Agents*, The University of Michigan Press, Michigan.
- Clark A. (2008), *Supersizing the Mind. Embodiment, Action and Cognitive Extension*, Oxford University Press, Oxford.
- Clark A. (2010), "Coupling, Constitution and the Cognitive Kind: A Reply to Adams and Aizawa," [in:] *The Extended Mind*, R. Menary (ed.), The MIT Press, Cambridge: 81-100.
- Clark A., Chalmers D. (1998), "The Extended Mind," *Analysis* 58 (1): 7-19.
- Cummins F. (2014), "Agency is Distinct from Autonomy," *Avant* V (2/2014): 98-112.
- Davidson D. (1978), "Intending," *Philosophy of History and Action* 11: 41-60.
- Davidson D. (1982), "Rational Animals," *Dialectica* 36 (4): 317-327.
- De Jaegher H., Di Paolo E. (2007), "Participatory Sense-Making: An Enactive Approach to Social Cognition," *Phenomenology and the cognitive sciences* 6: 485-507.
- Dennett D. (1976), "Conditions of Personhood," [in:] *The Identities of Persons*, A. Oksenberg Rorty (ed.), University of California Press, Berkeley: 175-196.
- Di Paolo E. (2009), "Extended Life," *Topoi* 28 (1): 9-21.
- Di Paolo E., Thompson E. (2014), "The Enactive Approach," [in:] *The Routledge Handbook of Embodied Cognition*, L. Shapiro (ed.), Routledge, New York: 68-78.
- Dworkin G. (1988), *The Theory and Practice of Autonomy*, Cambridge University Press, Cambridge.
- Elder-Vass D. (2010), *The Causal Power Or Social Structures. Emergence, Structure and Agency*, Cambridge University Press, New York.
- Elder-Vass D. (2014), "Social Entities and the Basis of Their Powers", [in:] *Rethinking the Individualism-Holism Debate. Essays in the Philosophy of Social Science*, J. Zahle, F. Colin (eds.), Springer, Cham: 39-53.
- Floridi L., Sanders J.W. (2004), "On the Morality of Artificial Agents," *Minds and Machines* 14: 349-379.
- Frankfurt H. (1971), "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68: 5-20.
- Froese T., Ziemke T. (2009), "Enactive Artificial Intelligence: Investigating the Systemic Organization of Life and Mind," *Journal of Artificial Intelligence* 173: 466-500.
- Goldspink Ch., Kay R. (2003), "Organizations as Self-Organizing and Sustaining Systems: A Complex and Autopoietic Systems Perspective," *International Journal of General Systems* 32 (5): 459-474.
- Heersmink R. (2012), "Mind and Artifact: A Multidimensional Matrix for Exploring Cognition-Artifact Relations," [in:] *Proceedings of the 5th AISB Symposium on Computing and Philosophy*, M. Bishop, Y. Erden (eds.), Bath: 54-61.

- Jonas H. (1968), "Biological Foundations of Individuality," *International Philosophical Quarterly* 8: 231-251.
- Kyselo M. (2014), "The Body Social: Enactive Approach to the Self," *Frontiers in Psychology* 5: 1-16.
- Luhmann N. (1986), "The Autopoiesis of Social Systems," [in:] *Sociocybernetic Paradoxes*. F. Geyer, J. van der Zouwen (eds.), SAGE Publications, London: 172-192.
- Luhmann N. (1995), *Social Systems*, Stanford University Press, Stanford.
- Markosian N. (1999), "A Compatibilist Version of The Theory of Agent Causation," *Pacific philosophical quarterly* 80: 257-277.
- Maturana H. (1987), "The Biological Foundations of Self-Consciousness and the Physical Domain of Existence," [in:] *Physics of Cognitive Processes*, E. Caianiello (ed.), World Scientific, Singapore: 324-379.
- Maturana H., Varela F. (1980), *Autopoiesis and Cognition – the Realization of Human Understanding*, Shambhala, Boston.
- Maturana H., Varela F. (1992), *The Tree of Knowledge – The Biological Roots of Human Understanding*, Shambhala, Boston.
- Mingers J. (2002), "Can Social System Be Autopoietic? Assessing Luhmann's Social Theory," *Sociological Review* 50 (2): 278-299.
- Moreno A., Etxeberria A., Umerez J. (2008), "The Autonomy of Biological Individuals and Artificial Models," *BioSystems* 91: 309-319.
- Pattee H.H. (2001), "The Physics of Symbols: Bridging the Epistemic Cut," *BioSystems* 60 (1/3): 5-21.
- Rohde M., Steward J. (2008), "Ascriptional and 'Genuine' Autonomy," *BioSystems* 91 (2): 424-433.
- Russell S., Norvig P. (1995), *Artificial Intelligence: A Modern Approach*, Prentice Hall, Englewood Cliffs.
- Sanches de Oliveira G., van Es T., Hipólito I. (2023), "Scientific Practice as Ecological-Enactive Co-Construction," *Synthese* 202 (1): 1-33.
- Thompson E. (2007), *Mind in Life: Biology, Phenomenology and the Sciences of Mind*, Harvard University Press, Cambridge.
- Thompson E., Stapleton M. (2009), "Making Sense of Sense-Making: Reflections on Enactive and Extended Mind Theories," *Topoi* 28 (1): 23-30.
- Varela F. (1979), *Principles of Biological Autonomy*, North Holland Publishing, New York.
- Varela F. (1991), "Organism: A Meshwork of Selfless Selves," [in:] *Organisms and the Origins of Self*, A. I. Tauber (ed.), Kluwer Academic Publishers, Dordrecht: 79-107.
- Varela F., Thompson E., Rosch E. (1991), *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press, Cambridge.