Reformalizing the notion of autonomy as closure through category theory as an arrow-first mathematics

Ryuzo Hirota[1], Hayato Saigo[2], and Shigeru Taguchi[3]

[1] Graduate School of Arts and Sciences, University of Tokyo
[2] Nagahama Institute of Bio-Science and Technology
[3] Center for Human Nature, Artificial Intelligence, and Neuroscience (CHAIN),
Faculty of Humanities and Human Sciences, Hokkaido University
hirota@sacral.c.u-tokyo.ac.jp

Abstract

Life continuously changes its own components and states at each moment through interaction with the external world, while maintaining its own individuality in a cyclical manner. Such a property, known as "autonomy," has been formulated using the mathematical concept of "closure." We introduce a branch of mathematics called "category theory" as an "arrow-first" mathematics, which sees everything as an "arrow," and use it to provide a more comprehensive and concise formalization of the notion of autonomy. More specifically, the concept of "monoid," a category that has only one object, is used to formalize in a simpler and more fundamental way the structure that has been formalized as "operational closure." By doing so, we show that category theory is a framework or "tool of thinking" that frees us from the habits of thinking to which we are prone and allows us to discuss things formally from a more dynamic perspective, and that it should also contribute to our understanding of living systems.

Introduction

Life continuously changes its own components and states at each moment through interaction with the external world, while maintaining its own individuality in a cyclical manner. Constructing an artificial system with such a property, i.e., "autonomy," has been one of the major challenges in artificial life (Aguilar et al., 2014, p. 5).

Mathematical formalization plays a major role in properly capturing such a property. A branch of mathematics called category theory (Mac Lane, 1971; Awodey, 2010; Simmons, 2011), which we will introduce later, has been used on occasion in theoretical biology as a formal framework that allows describing relations without relying on concrete components (e.g., Rosen, 1991; Varela, 1979; Nomura, 2007). And in recent years, there have again been active attempts to construct a system theory based on category theory (e.g., Capucci et al., 2022; Virgo et al., 2021; Fong and Spivak, 2019).

In particular, Francisco Varela, who with Humberto Maturana proposed the concept of autopoiesis (Maturana and Varela, 1980, 1987), also attempted to formalize his theory using concepts of category theory (Varela, 1979; Kauffman, 2017). These attempts, however, were eventually regarded as static and were gradually replaced by modeling with dynamical systems (e.g., Kelso, 1995; Thelen and Smith, 1994; Di Paolo et al., 2017), which are supposed to capture dynamic changes in the system. As a result, they have not gained as much attention from the current generation of researchers.

However, category theory can be interpreted as a step forward in the direction of grasping a very dynamic way of being. Although category theory is widely used by modern mathematicians, it is sometimes perceived as a static theory in that it describes the "invariant" structure behind various branches of mathematics. However, if we examine the original intentions of the theory's originators, it becomes clear that category theory was initially developed as a means of capturing the dynamic "movement" inherent in mathematical thinking[1]. In fact, the originators positioned their theory as an extension of the Klein's "Erlangen Program" (Klein, 1893) that sought to view geometry as a field centered on dynamic transformations, rather than static shapes to be transformed: "[Category theory] may be regarded as a continuation of the Klein Erlangen Program, in the sense that a geometrical space with its group of transformations is generalized to a category with its algebra of mappings." (Eilenberg and Maclane, 1945, p. 237)

Such an intention of the originators is deeply and subtly incorporated into the designs of the basic concepts of category theory. In particular, if we look carefully at the "axioms of category," the very basis of category theory, it becomes clear that they are rooted in a perspective that views everything not as a set of point-like elements, but as an "arrow."

[1] Some readers may think that category theory is static because it does not appear to explicitly incorporate temporal changes, as in dynamical systems. However, as will be mentioned later, (discrete) dynamical systems can be viewed as functors from a monoid to the category of sets (Spivak, 2014, p. 329), and from that perspective, the iteration of time in dynamical systems is nothing other than the composition of the arrows of the monoid. In other words, if dynamical systems are said to be dynamic, then so is a category (especially a monoid). There are also proposed models that allow the structure of a category to change dynamically and stochastically (e.g., Fuyama et al., 2020).

Based on this view on category theory, we use category theory not as a general-purpose language that can be applied to the mathematical modeling of general phenomena, but as a "tool of thinking" that leads us to a way of thinking that captures the dynamic nature of things by enabling us to avoid the "habits of thinking" to which we are prone.

As the first small step in this larger endeavor, we attempt here to reformulate a structure called "closure" using category theoretic thinking. It refers to the maintenance of certain characteristics of something constantly changing, and has been adopted by many researchers as a key concept in capturing the autonomous nature of life. One of the examples is the notion of "operational closure," originally proposed by Maturana and Varela (1980) and subsequently refined by Di Paolo and Thompson (2014). It describes how the processes that constitute a system are closed with respect to "enabling relation" and, together with the "precariousness" of a network of such processes, is claimed to define the autonomy of living systems and others.

While we believe that this concept is surely effective in capturing the fundamental characteristics of living systems, we also believe that there are some limitations with it.

Given this, from the perspective of category theory as an "arrow-first mathematics," we attempt to shed new light on the autonomous nature of life, which is described as "closure," by means of a structure known as a "monoid," which will be introduced in detail in a later section.

Although the concepts we use in this paper are mathematically very elementary, they nevertheless allow us to formulate in a concise and productive way the discussion on the autonomy of life that has so far been presented in an advanced natural language (or in other mathematical frameworks such as dynamical systems). This shows that category theory is a promising approach for studying complex and dynamic systems, including living organisms.

The structure of the paper is as follows. The next section provides an introduction to the basic concepts of category theory and shows that it incorporates an "arrow-first" perspective. The third section discusses the formalization of the autonomy of life as "closure," focusing on the concept of "operational closure" formally defined by (Di Paolo and Thompson, 2014). The fourth section introduces the concept of "monoid" (a category with a single object) as the basic mathematical concept to formalize the notion of operational closure in living systems, with a particular focus on their "self-mediating" nature. The discussion will examine the benefits and future prospects of formalization by category theory more generally. Finally, the conclusion will provide a brief summary of the paper's main findings.

Category theory as an "arrow-first" mathematics

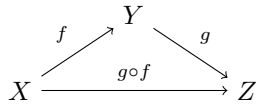
In this section, we introduce the concept of "category" and show that it has an "arrow-first" perspective.

Definition 1 (Category): A category is composed of two kinds of entities, namely, "objects" and "arrows (or morphisms)", that satisfy the following axioms. Any entities and relations that satisfy the axioms can be considered as "objects" and "arrows," respectively, regardless of their specific components.

Axiom 1 (Arrows and objects): Each arrow f has its "domain" (source) object $dom(f)$ and "codomain" (target) object $cod(f)$. An arrow such that $dom(f) = X$ and $cod(f) = Y$ can be expressed as $f : X \to Y$ or as follows:

$$X \xrightarrow{f} Y \tag{1}$$

Axiom 2 (Composition): A pair of arrows f, g can be "composed" into $g \circ f$ if the domain of one arrow is equal to the codomain of another, i.e., $cod(f) = dom(g)$. Assuming $f : X \to Y$ and $g : Y \to Z$, the following can be expressed:

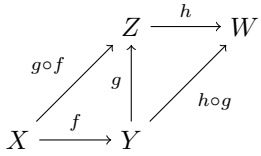


(2)

Axiom 3 (Associative law): The composition of arrows satisfies the "associative law," i.e.,

$$(h \circ g) \circ f = h \circ (g \circ f) \tag{3}$$

This means that assuming $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$, the following diagram is "commutative," i.e., no matter which path the arrows are composed through, if the start and end points are the same, the result is the same:



(4)

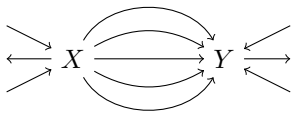
Axiom 4 (Identity): Each object X has its corrresponding arrow to itself $1_X : X \to X$ called "identity" such that for any arrow $f : X \to Y$,

$$f \circ 1_X = 1_Y \circ f = f \tag{5}$$

Intuitively, we can think of an object as representing a "thing," "phenomenon," or "event," and an arrow as representing a directed "relation," "process," or "transformation" between them. The most typical example that appears in the natural sciences is a category in which the distinct states of a

system are considered as the objects and possible transitions between the states as the arrows (Saigo et al. (2019) name this category the "category of mobility").

According to the basic idea of category theory, it does not begin by assuming objects unrelated to arrows; rather, an object is characterized only by what kind of arrows it has to other objects (and to itself). In this context, it is important to note that arrows are not reduced to pairs of objects. In general, there can be more than one arrow with the same domain and codomain, as schematically depicted in the diagram below:



(6)

Therefore, if we obtain only information about the objects from a category, we cannot recover the original category, whereas if we are given information about the arrows, we can reconstruct a complete picture of the original category. Unlike the set-theoretic approach, which typically first considers objects (sets) and then discusses the relations (functions) between them, it is precisely arrows that play the leading role in category theory.

Also important is the "identity law," which allows us to identify an object with its corresponding "identity" arrow and thus to regard an object as a kind of arrow. In other words, since a category consists of objects and arrows, and the objects are also a kind of arrow, we can say that a category is actually composed of various kinds of arrows. In this sense, again, "It's the arrows that really matter!" (Awodey, 2010, p. 8)

Based on these considerations, we can interpret category theory as an *"arrow-first"* mathematics, where everything is conceived of as an arrow. In this perspective, objects, which may appear static and fixed, are considered a particular type of arrow that are dedicated to *mediating* between arrows. Thus, category theory as "arrow-first mathematics" considers everything as an "arrow," while acknowledging the need for the identity arrows, which correspond to the objects. As a result, category theory does not merely reduce everything to transitive relations but rather always includes something individual and intransitive. This makes category theory different from simplistic relationalism or relational monism, which dissolve individuals into relations.[2]

[2]This stance of category theory is closely aligned with the one expressed by Varela (1976, 1979) in his phrase "Not one, not two." According to it, the dyad of "the it" and "the process leading to it" can be understood neither by separating them dualistically nor by reducing one to the other monistically; they can only be understood as complementary. Moreover, it is emphasized that they are not in a symmetrical relationship where the two are mutually exclusive, but in an asymmetrical relationship where one emerges from the other. Although it is beyond the scope of this paper, Varela analyzed such a complementary relationship using the concept of "adjunction" in category theory (Varela, 1979, pp. 97-99).

Furthermore, as mentioned above, there can be many (sometimes infinite) arrows between two objects in general (as depicted in the diagram (6)). Typically, we tend to think of a "relation" as being uniquely determined by the pair of objects related. However, in category theory, arrows are not reduced to pairs of objects. Therefore, the relations represented by arrows are not reduced to elements but are inherently diverse and pluralistic. These relations are more flexible than what we typically associate with the term "relation," and it may be more appropriate to use the term "mediation," which we will discuss in detail later.

We can also consider a category whose objects are arrows in another category, as we will discuss later. In that sense, what constitutes an arrow and an object is by no means predetermined. Rather, it depends on the context in which they are being used.

In summary, category theory characterizes mathematical objects not by what they are composed of, but by how they behave and relate to other things, i.e., what kind of arrows they have to and from other objects. Moreover, it views even the objects themselves as a form of arrow or process of "being an object," considering it an indispensable aspect. This nature of category theory would be compatible with artificial life, which seeks to characterize life by its relational and behavioral properties rather than by specific components such as proteins or DNA, but also to implement it as an individual.

In the following sections, we will consider the characteristics of living systems through the lens of category theory as an "arrow-first mathematics," which allows us to grasp the dynamic nature of things, as described in this section. In particular, our focus will be on the concept of "closure," which we will formalize using a special case of category called "monoid."

Autonomy and Closure

In theoretical biology, the notion of "closure" has often been used to capture the distinctive characteristics of life, particularly its autonomous nature (Moreno and Mossio, 2015). Simply put, the notion of closure refers to the phenomenon of returning to the original state after some manipulations, processes, or changes in a specific sense, i.e., being "closed" with respect to them. The notion of closure has been considered in various forms in theoretical biology. For instance, Rosen (1991) pointed out that the metabolic processes of life are "closed" with respect to "efficient causation" and attempted to formalize them using category theory (see also Letelier et al., 2003). Furthermore, Montévil and Mossio (2015) argue that the constraints on processes in living systems are closed: a process constrained by the outcome of

another process creates a constraint on yet another process, forming a cycle of constraints.

Here, we particularly focus on what is known as "operational closure." It basically means that some processes are "closed" with respect to their operations, and was originally proposed by Maturana and Varela (1980) as a property of "autopoietic" systems. It was also developed by Varela (1979) to characterize autonomous systems in general, not limited to biochemical processes in cells (including nervous, immune, or social systems). Subsequently, Di Paolo and Thompson (2014) define it more formally as the property that every process constituting a system is enabled by at least one other constituting process and also enables at least one other constituting process. In other words, it refers to how the processes that constitute the system are "closed" with respect to "enabling relations."

The concept of operational closure has become a foundation for the "enactive approach" in cognitive science, which argues that autonomy is at the basis of cognition. In contrast to mainstream cognitive science's computational and representationalist premises, the enactive approach, proposed by Varela et al. (1991), highlights that the agent and the world are not independently given but are brought forth or "enacted" through their interactions (Thompson, 2007; Di Paolo et al., 2017, 2018). The enactive approach is built on the theoretical traditions of autopoiesis and autonomy, but it also attempts to modify and expand them in fundamental ways (Weber and Varela, 2002; Di Paolo, 2005, 2009; Froese and Stewart, 2010). The redefinition of operational closure proposed by Di Paolo and Thompson (2014) is part of this effort to refine and extend the theoretical framework of autopoiesis and autonomy.

Figure 1 represents the concept of operational closure (Di Paolo and Thompson, 2014) . The diagram shows a series of circles that represent processes, and arrows that represent enabling relations between processes. When one process cannot occur without another, there is an enabling relation from the latter to the former (De Jaegher et al., 2010). The circles in black represent processes that are closed with respect to enabling relations, meaning that each black circle is enabling and enabled by another process represented by a black circle. The processes represented by these black circles, taken together, constitute an operationally closed system.

Typical examples of operationally closed processes are "self-distinction" and "self-production" in a living cell (Figure 2). The formation of a membrane creates a distinct physicochemical condition within it that facilitates metabolic reactions (self-distinction), while such metabolic reactions produce the components necessary for membrane formation (self-production), leading to a mutually enabling relation known as "autopoiesis," which was proposed by Maturana and Varela (Maturana and Varela, 1980, 1987; Varela, 1997; Di Paolo, 2018).

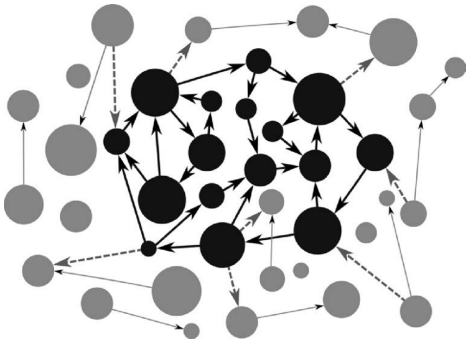


Figure 1: A schematic illustration of the concept of operational closure (reprinted from Di Paolo and Thompson, 2014). Copyright Ezequiel Di Paolo, 2013. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

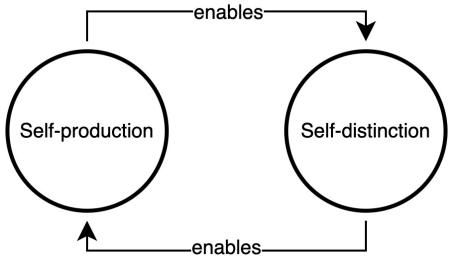


Figure 2: Operational closure between self-production and self-distinction.

Importantly, Di Paolo and Thompson (2014) argue that for a system to be autonomous, it is not enough that it is operationally closed; it must also be additionally "precarious." Precariousness means that the operationally closed processes are not given, but are non-trivial, produced and maintained by themselves. It is formally defined as the property that if any of the enabling relations cease to hold, the entire system cannot exist (Di Paolo and Thompson, 2014).

While the concept of operational closure is surely effective in capturing the autonomous nature of living systems, we consider that there are also some limitations with it. First of all, the enabling relation between processes (which are themselves a kind of relation) is so abstract and intangible that we cannot easily imagine, and as a result of trying to visualize such a hard-to-imagine thing, the visualization can easily cause a kind of reification of those processes. Of course, the authors emphasize that what are represented by the circles in the figure are not thing-like entities but processes, however, in our natural way of thinking, it is hard to imagine them properly, let alone their relationships. Therefore, there is a need for a concrete framework that allows for a more tangible understanding of these abstract relation-

ships, other than arguments in natural language and/or illustrations, as Beer (2020) also points out. As noted by Di Paolo et al. (2022), concepts in theories of autopoiesis and enaction are often misinterpreted and misapplied by researchers from outside the field, and we believe that this is partly due to the abstract nature of these concepts.

Furthermore, the figure could induce a way of thinking in which the processes depicted by the circles are viewed as spatially and/or functionally differentiated and modularized, similar to the schematic diagram of computational architecture. While this is not the intended interpretation of the authors, it is evident that the figure could be associated with functionalist and computationalist perspectives, which enactivists are supposed to be avoiding (Di Paolo et al., 2017, 2018).

To address these challenges, in the next section, we will attempt to reformulate the concept from a more fundamental and simple standpoint by using the language of category theory introduced in the previous section. Category theory is precisely a branch of mathematics that explores the properties of relationships between objects and makes it possible to concretize even highly advanced concepts, such as "relations of relations (of relations)," which are difficult (or almost impossible) to capture by natural language discussion alone. Our aim is not to refute the existing definition and depiction but to expand their potential by shedding light on them from a different perspective.

Also, the authors argue that "although the choice of processes under study is more or less arbitrary and subject to the observer's history, goals, tools, and methods, the topological property unraveled is not arbitrary" (Di Paolo and Thompson, 2014, p. 71), meaning that the definition's essence lies not in the specific processes represented by the circles (which vary depending on the observer) but rather in the properties of the enabling relations among them. As described in the last section, category theory is the mathematical framework created precisely to allow us to discuss such properties of relationships among objects without relying on their specific components. In this sense, too, our approach aligns with and complements the authors' intent.

Mediation and Monoid structure

For the purpose described in the last section, we will use the simplest mathematical formalization of the structure of what is called a "closure" in general: namely, a "monoid," a category with only one object. Importantly, even if it has only one object, there can be an infinite number of arrows (this follows from the irreducibility of an arrow to the pair of objects). This property allows us to successfully capture the way something is always varied, yet remains the same in a specific sense.

Definition 2 (Monoid): A monoid is a category with a single object.

Since it has only one object, a monoid is, in effect, a "collection of arrows." And since the domain and codomain of all the arrows are the same, it is possible to freely composite the arrows with each other. A monoid can be expressed as follows:

(7)

The concept of operational closure can be formulated as a monoid, insofar as it exhibits the characteristics of a "closure." In the following, we aim to achieve this by specifically focusing on "enabling relations" as arrows, following the basic stance of category theory as "arrow-first mathematics."

What is significant about an enabling relation is that it is not deterministic; "A enables B" does not mean that B will certainly occur or exist if A is present. Rather, what it exactly means is that *without A, there would be no B* (De Jaegher et al., 2010). This difference, often described as the one between "determination" and "dependence" (e.g., Di Paolo et al., 2018, p. 337), is crucial. Consider, for example, the relationship between seed and germination. Since germination is affected by various factors including environmental ones, it cannot be said that just because a seed exists, it will necessarily sprout. However, this does not mean there is no law and everything is uncertain. Rather, it is quite certain that there is a particular relationship: "If there is no seed, there will be no sprouting." This kind of relationship can be found everywhere in life phenomena in general. For example, one of the characteristics of life, at least on Earth today, is that individuals do not spontaneously arise, or simply put, "no children without parents" or "life only comes from life" (Oono, 2012; Froese and Taguchi, 2019). In this context, the relationship of "without A there would be no B" captures the nature of life as being inherently path-dependent and historical (see also Longo et al., 2012).

More generally, what has been studied as "causality" in biology, neuroscience, and medicine is essentially this "without A, there is no B" relation[3]: the relation between genetic "knockouts" and their effects on phenotype, the relation between physical and physiological changes in the nervous system and the transformation of subjective experience, and so on. Furthermore, a similar view of causality based on counterfactuals can also be found in the framework of statistical causal inference proposed by Pearl and colleagues (Pearl and Mackenzie, 2018), allowing for quantitative as well as qualitative analysis.

The relationship of "without A, there would be no B" can be conceptualized as one of "mediation" (Taguchi, 2019).

[3]Interestingly, such a conception of causality as a fundamental dependency on others can be closely related to what is called "dependent arising" (paṭicca samuppāda) in the Buddhist tradition, to which Varela et al. (1991) were also profoundly concerned.

This is exemplified by neural processes in the brain mediating human behavior or honeybees mediating the pollination of flowers, both of which imply that the former is necessary for the latter to occur. Moreover, the relationship "without A, there would be no B" does not exclude that it involves other factors than A and B. In essence, the relationship "without A, there would be no B" differs from the deterministic relationship "if there is A, there is always B" in that it typically involves other variables besides A that contribute to the occurrence or existence of B. For instance, the mere presence of a seed does not guarantee it will sprout; additional environmental factors such as soil, water, and heat are necessary.

Based on the notion of mediation as such, we can further obtain the perspective that all things are mediated by each other and things can exist only through such various forms of mediation. In other words, rather than something unmediated existing first and then entering into a mediating relationship, *mediation always comes first*, and what seems unmediated emerges only as a *mediating* point between mediating relations.

This mediation-based perspective is in deep accordance with the "arrow-first" perspective of category theory described above, in which objects do not exist independently as themselves but are characterized only as the "hubs" of arrows. In other words, the concept of category can accurately represent a worldview that sees everything as mediation.

From the perspective discussed so far, let us again consider the nature of living systems, especially the autonomy (and selfhood) expressed as operatinally closed, using category theory as a "tool of thinking".

In our lives, there are innumerable mediating relations among things, and some of them return to the original in a cyclic manner. They include horizontal relations with the outside, such as "agent $\rightarrow$ environment $\rightarrow$ agent" or "agent $\rightarrow$ other agents $\rightarrow$ agent," as well as vertical relations between the global and local inside the agent, such as "organism $\rightarrow$ organ $\rightarrow$ organism." They have been referred to as "circular (or reciprocal) causality" in the enactive approach and elsewhere (e.g., Thompson, 2007; Fuchs, 2017, 2020; Tschacher and Haken, 2007).

One typical example of this is chemotaxis. It refers to the tendency of microorganisms like E. coli to self-migrate towards environments richer in nutrients. In this case, the agent's existence as mobile allows for the presence of a specific environment around it through self-movement, and in turn, the environment enables the agent to survive. Thus, there is a mutually enabling and mediating relationship between the agent and the environment around it. In other cases, known as "niche construction" (Odling-Smee et al., 2003) or "extended physiology" (Turner, 2000), the agent "mediates" the environment more directly. As an example, Di Paolo (2009) describes how the water boatman, an aquatic insect, is able to keep air bubbles on its body surface underwater by means of hairs with water-repelling properties, which allows it to breathe and spend more time underwater, creating a mutually mediating relationship between the water boatman and its environment: water boatman $\rightarrow$ air bubbles $\rightarrow$ water boatman.

Considering a category whose arrows are the mediating relations between various objects, and then focusing on a single object and its various arrows from itself to itself (i.e., self-mediation), a monoid can be obtained (as a subcategory). As noted above, in category theory, relations represented by arrows are not reducible to the pairs of objects since the arrows between objects are more than one in general, as illustrated in diagram(6). In particular, there can be multiple or even innumerable arrows from an object to itself, with the "identity" arrow being the most trivial example. In the monoid considered here, the most trivial arrow, the "identity," is the tautological self-mediation "self $\rightarrow$ self," i.e., "without the self, there would be no self." However, as mentioned earlier, in category theory, an object itself does not have any characteristics and is characterized only in terms of its relations to others, i.e., arrows. This is also true of the monoid we are considering here, meaning that the self as the object does not exist independently of its relations to others, but can only exist as a "hub" through which the arrows are connected to each other. In other words, the "self" as the object of the monoid, or the self-mediation "self $\rightarrow$ self" as the identity arrow, can only exist through other forms of self-mediation, such as "self $\rightarrow$ environment $\rightarrow$ self," "self $\rightarrow$ other $\rightarrow$ self," or "self $\rightarrow$ organ $\rightarrow$ self."

Varela (1991) states that the autonomous self is "a meshwork of selfless selves" that is interwoven with various processes, including metabolic, immune, sensorimotor, and social ones, with no single substantial core. This conception of the self can be expressed in a natural way by the formalization of the self as a monoid described above.

Such a depiction of the autonomous self as a monoid might be interpreted as solipsistic, but this is not the case. In fact, the arrows of a monoid can be "factorized," which allows us to recover other factors as regularities that appear among the arrows. In the first place, an isolated factor cannot constitute mediation; mediation means that the existence of a factor intrinsically involves the existence of other factors. Therefore, the picture in which the self as an object can be grasped only in terms of its various mediations is thoroughly "world-involving" (Di Paolo et al., 2017).

Furthermore, category theory's "arrow-first" perspective stands in contrast to simplistic relational monism, which reduces everything into a web of relations and denies individuality. This aligns with the stance of the enactive approach, particularly of the one referred to as "autopoietic enactivism," which views the self as emerging from/through environmental interactions rather than as being a pregiven entity, yet without entirely dismissing the notion of an autonomous self as an illusion or observer's artifact (Baran-

diaran, 2017; Di Paolo, 2009), as expressed in the phrase "neither individualistic, nor interactionist" in the context of sociality (Di Paolo and De Jaegher, 2017). Thus, our view of autonomy as the monoid can also address the concern raised by Barandiaran (2017) that autonomy and the coupling with the environment are likely to be seen as a binary choice or tradeoff; What essentially constitutes the monoid are the arrows that represent self-enabling involving the environment, but these arrows are mediated by one object, namely the autonomous self, which in turn would be meaningless if isolated from the arrows.

One merit of using category theory as a language for formalization is that it allows for constructing a category from another category. One such construction is a "coslice category," whose object is an arrow in another category. More concretely, the objects and arrows of a coslice category A/Ω of a category Ω (where A is an object arbitrarily selected from all the objects of category Ω) are defined as follows:

Objects in A/Ω: Arrows in category Ω with A as its domain, such as $f : A \to X$ and $g : A \to Y$.

Arrows in A/Ω: An arrow in A/Ω between objects $f : A \to X$ and $g : A \to Y$ is defined as a triplet (f, g, t) such that an arrow $t : X \to Y$ in category Ω satisfies $g = t \circ f$, i.e., the following diagram is commutative in category Ω:

$$\begin{array}{ccc} X & \xrightarrow{t} & Y \\ & {\scriptstyle f}\nwarrow \quad \nearrow{\scriptstyle g} & \\ & A & \end{array} \tag{8}$$

Considering the coslice category of the monoid we have considered in this section, it is possible to take as objects the mediating relations that are the arrows in the monoid and to explicitly explore the relations between them (Figure 3). This is probably close to the figure of operational closure in Di Paolo and Thompson (2014) (Figure 1), in which each process is differentiated and depicted as an individualized circle.

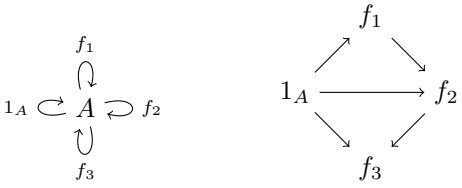


Figure 3: Monoid Ω (left) and Coslice category A/Ω (right).

For example, the mutual enabling relationship between self-distinction and self-production (illustrated in Figure 2) may be associated with the structure in the coslice category of the monoid.

The identity arrow 1_A of the object A in the original category Ω is included as well in the coslice category A/Ω as one of the objects and is placed along with other objects in the coslice category (i.e., arrows in the original category). However, this object is a special one known as the "initial object," which has an arrow to every object in the category, since any arrow $f : A \to X$ in the original category Ω and the identity arrow 1_A make the diagram below commutative and thus there is always an arrow in A/Ω from 1_A to f:

$$\begin{array}{ccc} A & \xrightarrow{f} & X \\ & {\scriptstyle 1_A}\nwarrow \quad \nearrow{\scriptstyle f\circ 1_A = f} & \\ & A & \end{array} \tag{9}$$

Such a property of an object possessing a unique arrow to each of all other objects in a category is known as the "universal property," and it plays a major role in the definition of various crucial concepts in category theory, such as those of "product" and "equalizer" (Leinster, 2014); These concepts are defined as the distinctiveness of a specific object in a category in terms of its relation with all other objects.

Back to the discussion on autonomy, the concept of universal property may also play an important role in characterizing the particularity of self-distinction in an autonomous system. According to Virgo et al. (2011), self-distinction, i.e., being a distinct unity (with spatial boundaries) is, on the one hand, nothing more than one of the constitutive processes (as depicted in Figure 2), but, on the other hand, it is still an exceptional one in that "it enables a great number of processes" (p. 247). In other words, self-distinction, although seemingly static, can be thought of as one, yet special, kind of process that mediates various, or perhaps even all, other self-producing processes. This would be depicted as a black circle with an arrow to all other black circles in a diagram as in Figure 1.

This feature of self-distinction may be associated with the universal property of the identity arrow 1_A (representing "being an object" as a process) of monoid Ω as the initial object of the coslice category A/Ω, possessing arrows to all other objects (Figure 3).This category-theoretic understanding of self-distinction, focusing on temporal and processual properties rather than spatial ones, can provide new insights into artificial life by generalizing the role of membranes.

Our intention in this section is not to argue that anything described as a monoid is autonomous or living. On the contrary, anything possessing a recursive property can be described as a monoid. This point would be partly related to the argument that autonomy requires not only being operationally closed but also being "precarious," that is, if any of the enabling relations between the constitutive processes disappear, the entire system ceases to exist (Di Paolo and Thompson, 2014). Such properties that distinguish the living from the non-living might be described as a specific interdependent structure in the coslice category, which needs to be addressed in future studies.

Discussion

Finally, in this section, we explicitly highlight some of the merits of employing category theory as a language for formalization.

First of all, category theory allows us to deal with the concept of "relation," which is so multifaceted that it often obscures the discussion, using a more precise concept: arrow. The word "relation" can sometimes be interpreted as reducible to the two terms it relates. In contrast, as discussed in this paper, arrows in category theory represent a specific and enriched notion, which we refer to as "mediation."

Another advantage of category theory as a language for formalization is that it allows us to associate between structures across different hierarchies. For example, it is possible to think of a part of an object (e.g., an organ of the self) or even what includes the object as a part (e.g., society to which the self belongs) as another object. In general, category theory can treat the part and the whole (or the local and the global) as being equal; they are equally treated as objects, and the relationship of containment between them is represented as an arrow between objects. Unlike the set-theoretic approach, which is based on and privileges the containment relationship "something is an element of another thing," category theory allows from the beginning to deal with relations in a broader sense as its default. This nature of category theory can make it possible to speak consistently about the hierarchical, vertical (local-global) relationship within the system and the horizontal relationship between the agent and the environment (including other agents).

In addition, although not discussed in this paper, category theory also allows us to rigorously describe higher-order relations, such as a correspondence between categories (an arrow in a category in which each object is a category), which is called a "functor," and even a consistent correspondence between functors (an arrow in a category in which each object is a functor), which is called a "natural transformation." It would be nearly impossible to speak of such higher-order relationships strictly in natural language alone. Category theory, in contrast, was designed by the originators from the beginning to address these higher-order relationships. As Leinster (2014) notes, "In fact, it was the desire to formalize the notion of natural transformation that led to the birth of category theory" (p. 9). And even in such categories as the "category of categories" and the "category of functors," the logic about autonomy discussed in this paper can hold true. Hence, these concepts could be applied to, for example, the formalization of the autonomy at the higher-order level discussed in the enactive approach, such as the autonomy of the sensory-motor schemes (Di Paolo et al., 2017) and the autonomy of social interactions (Di Paolo et al., 2018).

Finally, it should also be noted that dynamical systems, which have been frequently used in theoretical biology and cognitive science, including the enactive approach (e.g., Kelso, 1995; Thelen and Smith, 1994; Clark, 1998; Beer, 2004; Di Paolo et al., 2017) are one of the typical examples of monoids. The category of the "idea" of dynamical systems is a monoid whose arrows are generated by n-times compositions ($n = 1, 2, ...$) of a single arrow, and each discrete dynamical system can be regarded as a "set-valued functor" from this category to the category of sets (a category whose objects and arrows are sets and mappings between them). Thus, monoids can be seen as a relaxation of the deterministic property of dynamical systems, in which all arrows are generated from a single arrow. From our perspective, it is the structure of "repetition" that people have sought to capture using the language of dynamical systems, and the commitment to determinism is not necessarily essential. Indeed, the enactive approach is beginning to question the assumption of structural determinism (Froese and Taguchi, 2019; Froese, 2023; Fuchs, 2021) that was central to Maturana and Varela's original concept of autopoiesis (Maturana and Varela, 1980; Froese and Stewart, 2010). In this context, monoids, and category theory in general, potentially serve as a more suitable mathematical framework than dynamical systems.

Conclusion

In this paper, we have introduced category theory as an "arrow-first" mathematics to provide a comprehensive and concise formalization of the autonomous nature of life, especially the structure that has been expressed as "closure." Along the way, category theory has served not merely as a "neutral" formal language but as a tool for thinking that frees our minds from the habits into which they often fall, and leads us to view everything as a relation, or "mediation." This allowed us to formalize the notion of autonomy as a monoid closed under mediating relationships, and to see the autonomous self as a "hub" through which various self-mediating processes are mediated.

Artificial life has become one of the most interdisciplinary and hybrid fields, serving as the nexus of various fields such as biology, chemistry, artificial intelligence, robotics, cognitive science, and philosophy, which makes itself a very "mediating" field of research. Category theory shall serve as a promising platform in such a field for exchanging and integrating ideas from various areas in a comprehensive and productive manner and guiding our thinking in a freer direction.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Number JP20H00001 and JP23H04831. We would like to thank Tom Froese for his insightful discussion on an earlier version of this work and anonymous reviewers for their helpful comments.

References

Aguilar, W., SantamarÃa-Bonfil, G., Froese, T., and Gershenson, C. (2014). The past, present, and future of artificial life. *Frontiers in Robotics and AI*, 1:8.

Awodey, S. (2010). *Category theory*. Oxford Logic Guides. Oxford University Press.

Barandiaran, X. E. (2017). Autonomy and enactivism: Towards a theory of sensorimotor autonomous agency. *Topoi*, 36(3):409–430.

Beer, R. D. (2004). Autopoiesis and cognition in the game of life. *Artif. Life*, 10(3):309–326.

Beer, R. D. (2020). Lost in words. *Adapt. Behav.*, 28(1):19–21.

Capucci, M., Gavranović, B., Hedges, J., and Rischel, E. F. (2022). Towards foundations of categorical cybernetics. *Electron. Proc. Theor. Comput. Sci.*, 372:235–248.

Clark, A. (1998). *Being There: Putting Brain, Body, and World Together Again*. MIT Press.

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.*, 14(10):441–447.

Di Paolo, E. and De Jaegher, H. (2017). Neither individualistic, nor interactionist. In Durt, C., Fuchs, T., and Tewes, C., editors, *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*. MIT Press.

Di Paolo, E., Thompson, E., and Beer, R. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *PhiMiSci*, 3.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenol. Cognitive Sci.*, 4(4):429–452.

Di Paolo, E. A. (2009). Extended life. *Topoi*, 28:9–21.

Di Paolo, E. A. (2018). The enactive conception of life. In Newen, A., De Bruin, L., and Gallagher, S., editors, *The Oxford Handbook of 4E Cognition*, pages 70–94. Oxford University Press.

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

Di Paolo, E. A., Cuffari, E. C., and De Jaegher, H. (2018). *Linguistic Bodies: The Continuity between Life and Language*. MIT Press.

Di Paolo, E. A. and Thompson, E. (2014). The enactive approach. In Shapiro, L., editor, *The Routledge Handbook of Embodied Cognition*, pages 68–78. New York: Routledge.

Eilenberg, S. and Maclane, S. (1945). General theory of natural equivalences. *Transactions of the American Mathematical Society*, 58:231–294.

Fong, B. and Spivak, D. I. (2019). *An invitation to applied category theory: seven sketches in compositionality*. Cambridge University Press.

Froese, T. (2023). Irruption theory: A novel conceptualization of the enactive account of motivated activity. *Entropy*, 25(5):748.

Froese, T. and Stewart, J. (2010). Life after ashby: Ultrastability and the autopoietic foundations of biological autonomy. *Cybern. Hum. Knowing*, 17:7–49.

Froese, T. and Taguchi, S. (2019). The problem of meaning in AI and robotics: Still with us after all these years. *philosophies*, 4(14).

Fuchs, T. (2017). *Ecology of the Brain: The Phenomenology and Biology of the Embodied Mind*. Oxford University Press.

Fuchs, T. (2020). The circularity of the embodied mind. *Front. Psychol.*, 11:1707.

Fuchs, T. (2021). *In Defense of the Human Being: Foundational Questions of an Embodied Anthropology*. Oxford University Press.

Fuyama, M., Saigo, H., and Takahashi, T. (2020). A category theoretic approach to metaphor comprehension: Theory of indeterminate natural transformation. *Biosystems.*, 197:104213.

Kauffman, L. H. (2017). Mathematical work of francisco varela. *Constructivist Foundations*, 13:11–17.

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-organization of Brain and Behavior*. MIT Press.

Klein, F. (1893). Vergleichende betrachtungen über neuere geometrische forschungen. *Mathematische Annalen*, 43:63–100.

Leinster, T. (2014). *Basic Category Theory*. Cambridge University Press.

Letelier, J. C., Marín, G., and Mpodozis, J. (2003). Autopoietic and (M,R) systems. *J. Theor. Biol.*, 222(2):261–272.

Longo, G., Montévil, M., and Kauffman, S. (2012). No entailing laws, but enablement in the evolution of the biosphere. In *Proceedings of the 14th annual conference companion on Genetic and evolutionary computation*, GECCO '12, pages 1379–1392, New York, NY, USA. Association for Computing Machinery.

Mac Lane, S. (1971). *Categories for the Working Mathematician*. Springer.

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

Maturana, H. R. and Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. Shambhala, Boston, MA, US.

Montévil, M. and Mossio, M. (2015). Biological organisation as closure of constraints. *J. Theor. Biol.*, 372:179–191.

Moreno, A. and Mossio, M. (2015). *Biological Autonomy: A philosophical and theoretical enquiry*. History, Philosophy and Theory of the Life Sciences. Springer, New York, NY.

Nomura, T. (2007). Category theoretical distinction between autopoiesis and (M,R) systems. In *Advances in Artificial Life*, pages 465–474. Springer Berlin Heidelberg, Berlin, Heidelberg.

Odling-Smee, F. J., Laland, K. N., and Feldman, M. W. (2003). *Niche Construction: The Neglected Process in Evolution*. Princeton University Press.

Oono, Y. (2012). *The Nonlinear World: Conceptual Analysis and Phenomenology*. Springer Science & Business Media.

Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Rosen, R. (1991). *Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life*. Columbia University Press.

Saigo, H., Naruse, M., Okamura, K., Hori, H., and Ojima, I. (2019). Analysis of soft robotics based on the concept of category of mobility. *Complexity*, 2019.

Simmons, H. (2011). *An Introduction to Category Theory*. Cambridge University Press.

Spivak, D. I. (2014). *Category Theory for the Sciences*. MIT Press.

Taguchi, S. (2019). Mediation-Based phenomenology: Neither subjective nor objective. *Metodo Int. Stud. Phenomenol. Philos.*, 7(2):17–44.

Thelen, E. and Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press.

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

Tschacher, W. and Haken, H. (2007). Intentionality in non-equilibrium systems? the functional aspects of self-organized pattern formation. *New Ideas Psychol.*, 25(1):1–15.

Turner, J. S. (2000). *The Extended Organism: The Physiology of Animal-Built Structures*. Harvard University Press, Cambridge, MA.

Varela, F. J. (1976). Not one, not two. *CoEvolution Quarterly*, 12:62–67.

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

Varela, F. J. (1991). Organism: A meshwork of selfless selves. In Tauber, A. I., editor, *Organism and the Origins of Self*, pages 79–107. Springer Netherlands, Dordrecht.

Varela, F. J. (1997). Patterns of life: intertwining identity and cognition. *Brain Cogn.*, 34(1):72–87.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Virgo, N., Biehl, M., and McGregor, S. (2021). Interpreting dynamical systems as bayesian reasoners. *Joint European Conference on Machine*.

Virgo, N., Egbert, M. D., and Froese, T. (2011). The role of the spatial boundary in autopoiesis. In *Advances in Artificial Life. Darwin Meets von Neumann*, pages 240–247. Springer Berlin Heidelberg.

Weber, A. and Varela, F. J. (2002). Life after kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenol. Cognitive Sci.*, 1(2):97–125.