Autopoietic theory, enactivism, and their incommensurable marks of the cognitive

Mario Villalobos[1] · Simón Palacios[2,3]

Abstract

This paper examines a fundamental philosophical difference between two radical postcognitivist theories that are usually assumed to offer (more or less) the same view of cognition; namely the autopoietic theory (AT) and the enactive approach. The ways these two theories understand cognition, it is argued, are not compatible nor incompatible but rather incommensurable. The reason, so it is argued, is that while enactivism, following the traditional stance held by most of the cognitive theories, understands cognitive systems as constituting a (sort of) natural kind, the autopoietic theory understands them as constituting only a conventional kind. Additionally, the paper shows that AT's conventionalist stance about cognition, far from being an undesirable or useless position, offers some methodological virtues that might be timely and welcome in the agitated and revolutionary climate of current cognitive science.

Keywords Autopoietic theory · Enactivism · The mark of the cognitive · Conventional kind · Pluralism · Strict naturalism

1 Introduction

This paper will bring to light and examine a fundamental philosophical difference between two radical post-cognitivist theories that—perhaps due to their tight historical connection—are usually assumed to offer (more or less) the same view of

✉ Mario Villalobos
    mario.kirmayr@gmail.com

    Simón Palacios
    palaciosbriffault@gmail.com

[1] Escuela de Psicología y Filosofía, Universidad de Tarapacá, 18 de septiembre 2222, 1010069 Arica, Arica y Parinacota, Chile

[2] Instituto de Filosofía y Ciencias de la Complejidad, Los Alerces 3024, 7780192 Ñuñoa, Santiago, Chile

[3] Universidad Alberto Hurtado, Santiago, Chile

🌲 Springer

cognition; namely the autopoietic theory (AT) and the enactive approach.[1] These theories are historically linked, the latter being largely a development and transformation of the former. As is usual among family members, the theories share some basic assumptions, but also exhibit some important incompatibilities. Both theories agree that the behavior of living beings (even those that have brains) shouldn't be explained by appealing to internal representations and information-processing mechanisms, but they disagree about the more basic nature of living beings. The various areas of compatibility and incompatibility between the enactive approach and AT have already been examined in some detail by both enactivists and autopoieticists (Di Paolo 2005; Froese and Stewart 2010; Villalobos 2013; Villalobos and Ward 2015). However, there is a deeper gap that separates these theories which has not been yet analyzed in the literature, namely their incommensurable conceptions of cognition.

In this paper, we argue that the way AT understands cognition is neither compatible nor incompatible with the way enactivism understands it. Rather, the autopoietic view of cognition, so we will try to show, is incommensurable with the enactive one. The reason is that while enactivism, following the traditional stance held by most theories of cognition, understands cognitive systems as constituting a (sort of) natural kind, the autopoietic theory understands them as constituting only a conventional kind. We also argue that AT's conventionalist stance, far from being an undesirable or useless position, offers some methodological virtues that might be timely and welcome in the agitated and revolutionary climate of current cognitive science.

To demonstrate the incommensurability that (we argue) exists between the enactive approach and AT's views of cognition, and also to better understand the philosophical motivations behind AT's conventionalist stance, we will present a debate between both theories regarding the nature of living beings and the concept of cognition. Firstly, in Sect. 2, we will present what might be called the 'enactive mark of the cognitive', which, as we shall see, proves to be extensionally equivalent to the category 'living beings'. Secondly, in Sect. 3, we will introduce the autopoietic conceptions of living beings and of cognition, which, grounded in what we identify as a 'Strict Naturalistic' methodological commitment, provides a direct objection to the enactive conception. Then, in Sect. 4, we will construct a plausible reply on behalf of the enactivists, which asks AT to reconsider its objection on pain of sliding towards a dark and seemingly unavoidable dilemma between eliminativism and pan-cognitivism. Having reached this point, we will introduce AT's conventionalist stance about cognition, which allows AT to escape the dilemma while keeping intact its ontological and methodological commitments. Finally, by analyzing some potential objections, we will discuss the reach of AT's conventionalism and the potential pluralistic virtues it might offer in the current climate of cognitive science.

---

[1] The kind of enactivism that we have in mind in this paper is the canonical version developed by Varela and collaborators (Varela et al. 1991; Weber and Varela 2002; Thompson 2007; Di Paolo 2005; Froese and Stewart 2010), sometimes dubbed "autopoietic enactivism", but perhaps better called "autonomist enactivism". Other versions of enactivism, such as the sensorimotor theory of O'Regan and Noë (2001), and the "radical" branch of Hutto and Myin (2013), will not be considered here.

2 The enactive mark of the cognitive and the cognitive status of living beings

What is cognition? What makes a system or a process a cognitive one? What is, ultimately, the mark of the cognitive? The enactivist answer to these questions is straightforward: "sense-making is the basic mark of the cognitive" (Thompson 2011, p. 211). According to the enactivists, in defining "*cognition* as *sense-making* […] we capture most (perhaps all) of our important intuitions about the term [cognition]" (Di Paolo 2009, p. 15. Original emphasis). What are those intuitions? They are, so the enactivists claim, that cognition "is normative, asymmetric (…), relational, (…), and it implies a self-constituted identity to which norms refer" (Di Paolo 2009, p. 15). Let us unpack these ideas.

Sense-making, according to the enactivists, is "behaviour or conduct in relation to significance, valence, and norms that the system itself brings forth or enacts on the basis of its autonomy" (Thompson 2011, p. 211). The canonical example to illustrate sense-making at its most basic level is bacteria swimming up through a gradient of sugar (Weber and Varela 2002; Thompson 2005, 2007). Sugar is a chemical substance in the environment, but its meaning as 'food' is given only by the sense-making activity of the bacteria as autonomous systems:

> [the] significance and valence of sugar are not intrinsic to the sugar molecules; they are relational features, tied to the bacteria as autonomous unities. Sugar has significance as food, but only in the milieu that the organism itself enacts through its autonomous dynamics. (Thompson and Stapleton 2009, p. 25)

It is because the bacteria value sugar as something 'good' for their needs that sugar takes on significance as 'food.' And it is because the bacteria have 'needs' to fulfill that they can value something as 'good' or 'bad.' These features, i.e., having needs and the ability to value things as good or bad, are distinctive of what enactivists call 'autonomous systems.' Autonomous systems, according to enactivists, are systems that, due to their particular ontology (which will be analyzed soon), set their own values and engage with their environment in a normative way (Di Paolo 2009). This last point is so central for enactivists that, in fact, it is equally valid to say that the "normative engagement [of autonomous systems] is the hallmark of cognition" (Di Paolo 2009, p. 15). For enactivists, only autonomous systems are sense-making systems, and therefore, cognitive systems (Thompson 2011). Thus, in the example above, it is ultimately because bacteria are autonomous systems that they are cognitive (sense-making) systems.

According to the enactivists, an autonomous system is, essentially, "an *operationally closed* and *precarious* system" (Di Paolo and Thompson 2014, p. 69. Original emphasis). The exact characterization of these two defining features, operational closure and precariousness, is rather complex but is also not necessary for our purposes here (although see Villalobos and Dewhurst 2018 for a detailed critical analysis). The crucial point is that, for the enactivists, these are objective ontological features whose presence allows us "to answer empirically

the question of whether a system is autonomous" (Di Paolo and Thompson 2014, p. 72). For the enactivists, autonomous systems form (something close to) a natural kind, in the sense that their distinctive features are ontologically objective, and therefore empirically and operationally recognizable. As Di Paolo and Thompson claim, "[t]he enactive concept of autonomy is entirely operational, and therefore *naturalistic*" (2014, p. 72. Emphasis added).

The importance of this is that autonomy, according to the enactivists, is the key feature that allows us to distinguish cognitive systems from non-cognitive systems:

> the principal concept that differentiates enactivism from other embodied approaches to the mind [is] the concept of *autonomy*. By making use of this concept (…) we can give *operational criteria for distinguishing cognitive systems from non-cognitive systems.* (Di Paolo and Thompson 2014, p. 69. Second emphasis added)

Autonomy, in the enactive view, is the natural (ontological) mark of cognitive systems. It also provides, presumably, a naturalistic account of what enactivists take to be key properties of cognitive systems, such as "immanent purposiveness", "intrinsic teleology", and the "self-generation of norms" (Di Paolo and Thompson 2014, pp. 71–72. See also Di Paolo 2005; Thompson 2007).

Now, if all that is so, the next question is: What kind of systems, in the real world, are genuinely autonomous, and therefore cognitive? What, in the real world, is the extension of the class of 'autonomous systems', and therefore of 'cognitive systems'? Here, the previous example of the bacteria gives us a clue. According to the enactivists, the only natural systems known hitherto that can be granted as genuinely autonomous are living beings. Enactivism conceives of living beings as autopoietic systems and (in fact contrary to the AT, as we will see) takes autopoiesis to be the paradigmatic, most fundamental case of autonomy (Thompson 2007, 2011; Varela et al. 1991; Varela 1979). That is why, for enactivists, "there is an intimate relation between being alive and being cognitive" (Di Paolo 2009, p. 13). It is because living beings are autonomous systems that exhibit what enactivists take to be the distinctive properties of cognitive systems; i.e., teleology, normativity, and agency.

Enactivists take autopoiesis to be a sufficient condition for autonomy (i.e., wherever there is life there is autonomy). They are not entirely sure, however, about whether autopoiesis should also be considered a necessary condition for autonomy. On the one hand, enactivists do not want to rule out, a priori, as a matter of simple conceptual definition, the possibility of autonomous systems that are not living systems. On the other hand, they admit that is hard to think of an autonomous system that is not at the same time something at least organizationally equivalent to an autopoietic system (or something made of autopoietic components). Thompson, for example, claims that "it is not unreasonable to doubt that [autonomy] can be achieved without autopoietic constituents" (2011, p. 215).

For all practical effects, considered as an empirically based research program, it is fair to say that enactivism takes the class of cognitive systems (i.e., real autonomous systems) to be coextensive with the class of living systems.

3 AT's conception of living beings and cognition

Enactivism, as we saw in the previous Section, defines cognition as sense-making, and ascribes the ability to sense-make to living beings, under the assumption that living beings are autonomous systems. AT, as we will now see, rejects the enactive conception of living beings as autonomous systems, and offers a different conception of cognition.

AT rejects the enactive conception of living beings, in short, because it violates what AT takes to be a fundamental and non-negotiable methodological assumption in the scientific practice: Strict Naturalism.

Briefly sketched, AT's Strict Naturalism is a methodological stance that rejects the inclusion, either in our conception or explanation of living beings, of properties which are not found in the theoretical practice of the natural sciences. According to AT's Strict Naturalism, living beings are natural systems and must be studied as such; that is, by appealing to the same ontological assumptions and explanatory principles that the current science uses to study any natural system in general. Roughly stated, these ontological assumptions are, among others, that natural phenomena are what they are and occur as they occur: (1) without having any goals, purposes or intentions (i.e., without teleology), and (2) without being governed by any consideration of what may be good or bad, correct or incorrect, adequate or inadequate, beneficial or harmful (i.e., without normativity).

AT defines living beings as physical autopoietic systems (Maturana 1975, 1981; Maturana and Varela 1980). In this definition, however, what matters for AT's Strict Naturalism is not that living beings are autopoietic but rather that they are physical (natural) systems (Maturana 2011; Villalobos 2013; Villalobos and Ward 2015). A crucial point in the theoretical agenda of AT is to show that once living beings are revealed as (being nothing more than) autopoietic systems, it becomes apparent and clear that they are trivially natural systems. AT's Strict Naturalistic reasoning is that if natural systems such as planets, stars, rivers, and volcanoes are not conceived of as having teleology and normativity, then, living beings, also being natural systems, should not be so conceived either.[2] According to AT, "living systems, as physical autopoietic systems, are purposeless systems" (Maturana and Varela 1980, p. 86), and "what [we] call *normative activities are not aspects of* [...] *autopoiesis* [but only] commentaries or explanatory propositions that [we] make about what [we] may think that should occur in the [...] organism" (Maturana 2011, pp. 149–150, Emphasis added).

Enactivism and AT have, thus, something like a 'deep disagreement' about the ontology of living beings. Both theories agree that living beings are autopoietic systems but draw entirely different conclusions from that. For enactivism, it is

---

[2] Notice, however, that this prescription does not come as an axiomatic or a priori judgment. AT's Strict Naturalism is a methodological stance that follows the lead of the natural sciences and that, therefore, remains open to be informed and updated by their progress and discoveries. If the natural sciences demonstrate, at some moment, that living beings do have teleology and normativity as natural properties, AT should update its conception of living beings accordingly. The point for the present discussion is that, lacking such a demonstration, AT's Strict Naturalism currently dictates the rejection of such properties.

because living beings are autopoietic systems that they are autonomous systems, and therefore, teleological and normative systems. For AT, on the contrary, it is precisely because living beings are autopoietic systems that they are revealed as non-teleological and non-normative systems; that is, as non-autonomous systems (in the enactive sense).

That is the fundamental disagreement between AT and enactivism about the nature of living beings. What about cognition? AT, in its early formulation, claimed that "living systems are cognitive systems," and that "*living as a process is a process of cognition*" (Maturana and Varela 1980, p. 13. Original emphasis). This formula seems to imply that life is necessary and sufficient for cognition, which is precisely the interpretation that the enactive approach, largely inspired by the AT, seems to take. However, paying attention to its more contemporary version, as we will see, it becomes clear that AT is not committed to the idea that only living beings should be assigned a cognitive status. Let us briefly reconstruct AT's rationale regarding the notion of cognition.

AT starts by considering (what it takes to be) our intuitive use of epistemic notions such as 'knowing' and 'cognition' in the domain of living beings, to then unpack in theoretical terms what is presumably behind said notions:

> [I]f we see a living system behaving according to what we consider is adequate behavior in the circumstances in which we observe it, we claim that it knows. What we see in such circumstances underlying the adequate behavior of the living systems is:
> a) that the living system under our attention shows or exhibits a structural dynamics that flows in congruence with the structural dynamics of the medium in which we see it; and,
> b) that it is through that dynamic structural congruence that the living system conserves its living. (Maturana 2002, p. 26)

According to AT, "we human beings call cognition the capacity that a living system exhibits of operating in dynamic structural congruence with [its] medium" (Maturana 2002, p. 26). This capacity of structural congruence has to do with a fundamental process of structural coupling or adaptation. AT calls "structural coupling or adaptation the relation of dynamic structural correspondence with the medium in which a unity conserves its class identity [i.e., its integrity]" (Maturana 2003, p. 64). The connection that AT sees between cognition (knowing), structural (or operational) congruence and structural coupling is summarized as follows:

> the process which gives rise to the operational congruence between an organism and its niche, that is the process that we distinguish in daily life as (…) knowing, is *structural coupling*. (Maturana 2002, p. 26. Original emphasis)

Living beings exhibit adaptation and structural coupling, and that is why we see them existing in structural congruence (correspondence) with their medium (i.e. their environment). This process of structural coupling or adaptation is, ultimately, what AT claims we identify with the concept of cognition.

When put this way cognition, for AT, seems to be something equivalent or very close to biological adaptation, and therefore (just as the enactivists assume) a capacity intimately linked to living beings. However, AT understands adaptation and structural coupling as capacities that are not restricted to the biological domain:

> If the organization of a composite unity [i.e., a system] remains invariant while it undergoes structural changes […] through its recurrent interactions in its medium [we say that] its adaptation is conserved […]. Defined in this manner, […] conservation of adaptation *is not peculiar to living systems*. It is a phenomenon that takes place whenever a plastic composite unity undergoes recurrent interactions with structural change but without loss of organization. (Maturana and Varela 1980, pp. xx-xxi. Emphasis added).

According to AT, *every system* in interaction with its medium, as long as it exhibits structural plasticity and conserves its integrity, is in structural coupling and adaptation. Living beings' structural coupling and adaptation is, thus, only one version of this general condition.

> Given that structural coupling in its domain of existence (conservation of adaptation) is a condition of existence for *any system* distinguished by an observer, (…) when an observer distinguishes a living system, he or she necessarily distinguishes it as a system that constitutively remains in structural coupling in its domain of existence. (Maturana 2003, p. 90. Emphasis added)

AT says that adaptation and structural coupling are, ultimately, the conditions or processes which we refer to when we talk about cognition, and it also says that the processes of adaptation and structural coupling are not exclusive to living beings. Why, then, do we restrict the notion of cognition only to living beings? Is there any key difference between the structural coupling and adaptation of living beings, as compared with those of other systems, such that only the former can be described as cognitive? Is there any principled reason to keep the concept of cognition attached exclusively to living beings, or is this just an arbitrary decision? Here is what Maturana, the main representative of AT, thinks about this point:

> I speak of cognition only in relation to living systems. [However, t]his is *arbitrary* since what I have said in relation to existence applies to *every entity* brought forth through an operation of distinction. (Maturana 2003, p. 95. Emphasis added)

It is clear, therefore, that in AT the concept of cognition denotes the process of structural coupling and adaptation that exists between an entity and its medium, regardless of whether the entity under consideration is or is not a living being. The choice to restrict cognition only to living beings is an arbitrary one, and not, as the enactivists claim, in some way principled or definitional of the concept as AT understands it.

4 AT's dilemma

In the previous Section, we have seen how AT departs from enactivism regarding the conception of living beings and cognition. For AT, enactivism is wrong in conceiving of living beings as autonomous systems; i.e., as endowed with properties such as normativity and teleology. In doing so, enactivism, according to AT, inadmissibly violates Strict Naturalism—as AT sees it, enactivists are guilty of appealing to spooky (i.e., non-scientifically credited) concepts. Enactivism is also wrong, according to AT, in ontologically attaching cognition only to living beings. For AT, the concept of cognition has to do with processes of structural coupling, which, strictly speaking, do not have any ontologically necessary dependence on life. In summary, despite the shared theoretical roots of AT and enactivism, for AT enactivism is wrong about almost everything when it comes to understanding living beings, cognition, and their relation.

The enactivist might want to reply here by pointing out some apparently unpalatable theoretical consequences of AT's strategy—problems that, as we will see, could take the form of a very unattractive dilemma. The enactivist will first point out that AT's conception of cognition as structural coupling and adaptation trivializes the concept of cognition to the point that practically everything becomes cognitive. If cognition is structural coupling, and if "structural coupling (…) is a condition of existence for any system distinguished by an observer" (Maturana 2003, p. 90), then all the physical systems we distinguish as observers (e.g., stars, planets, rivers, volcanoes, stones, etc.) become, by the mere fact of existing, cognitive systems. Likewise, if cognition is adaptation, and if adaptation "is not peculiar to living systems", but rather a phenomenon that applies to every structurally plastic system (Maturana and Varela 1980, pp. xx–xxi), then, given that practically all physical systems are able to undergo some structural changes without loss of integrity, almost everything becomes a cognitive system. AT's notion of cognition as structural coupling and adaptation, the enactivist will point out, leads us, in the best of cases, to an extremely liberal version of cognition, and at worst, to pan-cognitivism.

The enactivist, then, will argue that once we rule out structural coupling as a mark of the cognitive, on the grounds that it is too liberal, the only candidates we are left with are those that AT theorist rejects (i.e., sense-making, normativity, etc.), because they violate Strict Naturalism. Sense-making, recall, is "behaviour or conduct in relation to significance, valence, and norms that the system itself brings forth or enacts on the basis of its autonomy" (Thompson 2011, p. 211). Set in her non-negotiable Strict Naturalist commitment, AT's defender is forced to reject the possibility that properties such as normativity and autonomy may occur in natural systems, and therefore to conclude that cognition does not have a place in the natural world. This means, the enactivist will argue, that AT faces the dilemma of either endorsing pan-cognitivism (because it is too liberal) or eliminativism (because it is too strict and conservative).

According to the enactivist, AT's defender seems to face a quite dark dilemma. On the one hand, if she rejects the enactive mark of the cognitive and sticks with the idea of structural coupling, her fate is pan-cognitivism. On the other hand, if for

some remote reason she accepted the enactive mark of the cognitive, her fate would be eliminativism. Both options imply a deep theoretical devaluation of the concept of cognition, and are, therefore, highly unpalatable.

The only plausible combination, the enactivist will suggest, is to take both the enactive ontology (i.e., to drop Strict Naturalism and accept normativity as a natural property) *and* the enactive concept of cognition as sense-making. Any other combination is doomed to fail. Strict Naturalism plus cognition as structural coupling leads to pan-cognitivism. Enactive ontology plus cognition as structural coupling leads equally to pan-cognitivism. Strict Naturalism plus cognition as sense-making leads to eliminativism. The only alternative left for AT's defender, or so the enactivist will conclude, is to embrace full enactivism.

5 Escaping the dilemma: the idea of cognition as a conventional kind

The hypothetical enactivist reply presented in the previous Section seems to force AT's defender to embrace full enactivism, on pain of sliding down to a fatal dilemma between pan-cognitivism and eliminativism. However, as we will now see, AT's defender has independent reasons to reject the enactivist dilemma. The dilemma, according to the enactivist, arises because AT needs to choose between two different ways of fixing the natural mark of the cognitive: in terms of either structural coupling (adaptation) or sense-making. If cognition is structural coupling or adaptation, then practically every natural system is cognitive. If sense-making does not exist as a natural kind available to be deployed as a mark of the cognitive, because this violates Strict Naturalism, the only remaining option is eliminativism. Both structural coupling (adaptation) and sense-making are natural phenomena, and the extension of the concept of cognition corresponds either to one kind of phenomena or to the other. The presupposition here is that the scientific concept of cognition must pick out a natural process, condition or property, whose presence in the world should count as the truth-maker of our applications of the concept of cognition. That is to say, that 'cognition' as a scientific theoretical concept must correspond to a natural kind.

AT's defender, however, may object to this presupposition, and observe that the scientific concept of cognition, to begin with, does not need to correspond to a natural kind. The scientific concept of cognition, AT's defender will point out, maybe also be legitimately conceived of as a conventional kind; an option that the enactivist seems to have implicitly ruled out before initiating the debate.

If cognition is understood as a conventional kind, then, notice that the classification of some processes, properties or systems in the world as cognitive does not have any ontological import over said processes, properties or systems. If 'cognition' (or 'cognitive') is a concept we attribute to certain natural phenomena or systems by stipulation, that is to say, by a certain pragmatic or methodological convenience, then it does not make sense to search for a natural mark of the cognitive, as though this mark was an intrinsic property of certain processes or systems in the world. Under a conventionalist view of cognition, i.e., a view with no

ontological import, the enactivist dilemma, which runs at the ontological level, seems to lose its force (more on this soon).

Before evaluating the possible advantages and disadvantages that a conventionalist approach to cognition might have, let us see whether the approach is actually available to AT's defender. A closer look at the autopoietic literature shows that, indeed, the conventionalist view of cognition is a part of AT. Recall how AT starts the characterization of the concepts of knowing and cognition:

> [I]f we see a living system behaving *according to what we consider is adequate behavior* in the circumstances in which *we* observe it, *we claim* that it knows. (…) *[W]e human beings call cognition* the capacity that a living system exhibits of operating in dynamic structural congruence with [its] medium (Maturana 2002, p. 26).

AT's starting point is to consider the way that we observers qualify living beings as knowing, as being cognitive. The invitation is to see our use of the concept of cognition ("we human beings call cognition"), not to tell us what cognition is. As we saw in Sect. 3, AT argues that what we consider to be 'adequate behavior' is nothing but the structural congruence that living beings exhibit in relation to the medium in which we observe them. And it concludes that the (universal) process underlying said congruence is structural coupling (or adaptation). This whole characterization of cognition takes place, we must notice now, without abandoning the descriptive and distant stance taken as a starting point: "we human beings call this and that cognition". Why the use of this rhetoric formula? Why not just say, speaking as a theory, "cognition *is* this or that"? AT, if we reread Sect. 3, does not provide us with a theoretical definition of cognition, but rather with a clarification of the underlying processes that are indirectly connoted when we use the concept of cognition. The reason? AT assumes that 'cognition' is an attribution we make to certain systems or behaviors, not an intrinsic property or condition of said systems or behaviors. It is, so AT assumes, an ascription, not a description.

Commenting on the way some prominent enactivists (Froese and Stewart 2010) describe and explain living beings, Maturana says:

> The authors speak of adaptivity [in the normative enactive sense], goal directed actions and *cognition* as if they were referring to operational processes or properties of living systems, but those notions as presented do not connote biological processes but connote *opinions* of the speaker about the nature of what occurs with the living being in the flow of its living. (Maturana 2011, p. 150, footnote)

The concept of cognition, according to AT, expresses an opinion of the observer, not an intrinsic property of the observed system. More explicitly:

> [C]ognition is what an observer says that occurs when he or she sees an organism behaving in a manner that he or she considers adequate to the circumstances in which he or she observes it. (Maturana 2011, p. 150, footnote)

That is, the observer stipulates (implicitly or explicitly) some evaluative criteria of behavioural adequacy (i.e., "this behaviour is adequate, correct, successful, functional, intelligent," "whereas that other behaviour is inadequate, wrong, unsuccessful, dysfunctional, non-intelligent,"), and applies the concept of cognition, typically, to the cases she estimates as "adequate" or "intelligent". The concept of cognition, based on evaluative criteria, is not descriptive but rather normative, in the sense that it expresses not what the observer sees is the case, but rather what the observer sees is the case in relation to what she thinks should or ought to be the case:

> What the authors call normative activities such as adaptivity or goal directed actions or *cognition* (…) are not aspects of the dynamics of (…) autopoiesis, they are *commentaries or explanatory propositions* that an observer can make about what he or she may think that *should* occur in the flow of living of an organism in its relational domain. (Maturana 2011, p. 149–150)

If, under this light, we reread the characterization of cognition given by AT, we come to see that it is not that cognition *is* structural coupling or adaptation, but rather that these processes, when normatively interpreted as adequate (correct, intelligent) behavior, count as cognitive *for us*. 'Structural coupling' and 'adaptation' are theoretical concepts that denote natural processes; they constitute natural kind terms. 'Cognition,' according to AT, instead constitutes a conventional kind term (Villalobos and Silverman 2018).

AT's defender thus has ready to hand a conventionalist strategy to face the enactivist dilemma. The dilemma, recall, forces AT to take a position regarding the natural mark of the cognitive. If cognition is structural coupling, the consequence is pan-cognitivism. If Strict Naturalism rules out sense-making as a mark of the cognitive, then the consequence is eliminativism. Knowing that Strict Naturalism is a non-negotiable principle for AT, the rather perverse question asked by the enactivist is: "So, what is, according to you, the natural mark of the cognitive? Structural coupling (adaptation) or sense-making?" And the rather unexpected answer should be: "Neither, for there is no such a thing as a natural mark of the cognitive."

If the cognitive is conceived of as a conventional kind, then AT's defender is not forced to take any ontological commitment between structural coupling and sense-making as a definition of cognition. She might, of course, take a conventional commitment, as a matter of pragmatic (theoretical or methodological) convenience, but such a commitment, having no ontological import, cannot have any ontological consequences, whether pan-cognitivist or eliminativist. The enactivist dilemma is thus bypassed, and the Strict Naturalist commitment reaffirmed, since it is precisely because the concept of cognition, as the enactivist observes, "is normative" (Di Paolo 2009, p. 15), that, according to Strict Naturalism, it must be situated in the realm of our conventions, not in the realm of natural properties.

What if we decide (conventionally) to qualify every natural system as cognitive? Would that not lead us to pan-cognitivism? No, because such a decision, if it is a matter of convention, does not entail any ontological consequences about the nature of those systems. But if the concept of cognition is just a matter of convention, would that not lead us to eliminativism? No, because the conventionalist stance, being aware that non-natural kinds, including conventional kinds, play an important

and even indispensable role in reputable sciences (Pöyhönen 2013), does not promote the elimination of the concept of cognition.

Many well-established sciences, especially (though not exclusively) in the social field, have non-natural theoretical constructs (usually called 'social kinds'; cf. Guala 2014; Khalidi 2013) that satisfy reliable inductive inferences, grant fruitful explanations, and permit the normal development of empirical research programs. These sciences (e.g., sociology, anthropology, economics), though structured around conventional kind terms (e.g., class, gender, money), constitute respectable and useful empirical disciplines. A comparative example may be helpful here.

When a chemist sees a piece of gold, what she sees is a determinate molecular composition and a series of natural properties (atomic weight, melting point, conductivity). When an economist sees a piece of gold, what she sees is a certain amount of 'exchange value' or 'money.' The object of study of the chemist is, we might say, ontologically primary (gold: a natural kind) with respect to that of the economist (money: a conventional kind). This ontological hierarchy, however, does not render the economist's knowledge superfluous or less important than that of the chemist, or his discipline (economics) as less respectable or worthy of consideration than chemistry. If the economist is planning to spend a huge amount of money in buying a monumental statue made of gold, and the chemist reports to him that the statue is not really made of gold (i.e., that it is fake gold), the economist will surely take the report seriously and act accordingly. But likewise, if the chemist is planning to invest all her savings in gold mining companies in Asia, and the economist reports to her that, according to international trade projections the commercial wars between Asia and America will heavily devaluate those companies, the chemist will surely take this report seriously and act accordingly. She will do this even though the report has been elaborated, to a large extent, on the base of conventional kinds. The conventional kinds of the economist may not 'carve up *Nature* at its joints,' but that does not render his predictions worthless.

But even in the context of the natural sciences, there are examples of venerable theoretical terms which arguably do not correspond to natural kinds, and that are nevertheless kept around as part of scientific practice. One of the most notable cases is that of evolutionary groups concepts such as 'species' and 'clade' in biology. Some philosophers of biology, and some notable biologists too (cf. Darwin 1859), seem to agree that 'species' is not a natural kind term but rather a conventional one (Mishler 1999; Barker and Velasco 2013). However, none of them draws from that the conclusion that biologists should stop talking about 'species' altogether and eliminate the concept from their vocabulary. One reason for this is that the concept of 'species', even if conventional, and so perhaps open to a pluralist treatment, seems to play an important role as (what is called) an "investigative-kind concept" (Pöyhönen 2013; Brigandt 2003); i.e., a concept that usefully guides and organizes scientific research on the basis of a set of theoretical hypotheses (see Brigandt 2003, pp. 1308–1312).

The moral of these examples is that whereas AT's Strict Naturalism rules out the possibility of a natural kind type strategy for the identification of the mark of the cognitive, it does not rule out a scientifically useful, empirically informed and naturalistically-based conventionalist conception of cognition, nor does it promote an

eliminativist strategy. Cognitive science, as we will discuss in the next and last Section, has arisen and developed until now without having a natural kind definition of cognition, and it is not clear that it really needs one to move forward as a scientific discipline. On the contrary, as some observers will suggest, what the current agitated and revolutionary times of cognitive science seem to demand is not a(nother) sectarian war to establish an alleged "true and correct natural mark of the cognitive", but instead a pluralistic methodological framework to navigate the process of making the most of its theoretical diversity. If that is so, then the AT's defender would have a good reason to stick with her conventionalist stance, for conventionalism, as one might infer, naturally promotes pluralism about cognition.

6 Discussion

In the previous Sections we reviewed the enactive and autopoietic conceptions of cognition, and we have seen the way in which they radically differ. Whereas enactivism assumes cognition to be the (allegedly) natural process of sense-making and identifies cognitive systems with the (also allegedly) natural kind of autonomous systems, AT rejects the very idea that notions such as cognition or cognitive system pick out some sort of natural kind in the world. The enactive definition of cognition as sense-making is incompatible with alternative views that assert that cognition is, as a matter of natural fact, something else (e.g., information processing or the manipulation of internal representations). And it is compatible with views that assert that cognition is, again as a matter of natural fact, something very similar to sense-making (perhaps, for example, skillful engagement with affordances, as the sensorimotor enactivists propose). However, AT, in assuming a conventionalist approach to cognition, escapes this whole question of compatibility versus incompatibility, because it does not offer a competing definition of cognition as a natural kind. That is why we hold that the enactive and AT approaches to cognition are not directly compatible or incompatible, but rather incommensurable.

In this final Section we will discuss the reach and potential implications of AT's conventionalism, having as a background the current theoretical scenario of cognitive science. AT's conventionalism, as we saw in Sect. 5, seems to have the resources to escape the dilemma set by the enactivist, but, at what price? What might be the cost, for cognitive science, of assuming a conventionalist approach to cognition? Let us review two potential objections.[3]

---

[3] Here we address and discuss only some potential *methodological* implications of conventionalism in the field of cognitive science. There are, however, other and deeper implications of conventionalism, such as those that arise at the *metaphysical* level. One of the main worries at this latter level is that if cognition is taken to be something we merely ascribe to certain systems, then it is not easy to see how we could explain our own (and real) ability to ascribe cognition (or anything) in the first place. This is an interesting and important problem for any conventionalism about cognition, but one that unfortunately we cannot address here. See, however, Villalobos and Silverman (2018), and Abramova and Villalobos (2015), for some clues on how AT might deal with this kind of metaphysical worry.

Objection 1 (O1) Without a unique, natural kind-based definition of cognition, cognitive science does not have a definite object of study, and this suggests that it is just a degenerative scientific research program. Given that AT's conventionalism denies the possibility of finding a natural kind-based mark of the cognitive, it implies that cognitive science is on its way to methodological failure.

Reply We believe that O1 is not cogent, because it presupposes a far too strict methodological demand. Having a unique, natural kind-based definition of a scientific object of study is, perhaps, an ideal conceptual desideratum, but is not a realistic standard of good science. O1 assumes that only research programs that have clear, unique and natural kind-based definitions as their objects of study are progressive research programs. This standard of scientific status, however, would force the objector to qualify as degenerative, for example, one of the most successful and dynamic empirical sciences of our era, namely biology. Biology lacks a universal and conceptually neat definition of "life" (Tirard et al. 2010; Machery 2012; Cleland and Chyba 2002), but nothing important seems to hinge on such a condition. Biology has never needed, and probably does not need, a clear and unique natural kind-based definition of its object of study to develop as a research program, and it seems unjustified to demand of cognitive science a requirement that better-established disciplines, such as biology, do not meet or need (see Allen 2017, pp. 4244–4245).

AT's conventionalism denies the possibility of finding a natural kind-based definition of cognition, true, but that does not imply that cognitive science is degenerate. At any rate, whether cognitive science does or does not slide towards becoming a degenerative research program is something that, based on the example of biology, does not seem to depend on having or not having a single natural kind-based definition of its object of study. That is why perhaps philosophers and cognitive scientists of different stripes, all of them quite far away from AT's circles, have pointed out that lacking a universal or natural mark of the cognitive is not really something that is crucial for the future of the discipline.

Rupert (2013), for example, has given strong reasons for being pessimistic about the possibility of finding a natural kind that unifies our concept of cognition, but thinks that this situation doesn't motivate a general worry about the future development of cognitive science. Similarly, Allen, talking about the search for a natural mark of the cognitive, claims that "cognitive science will proceed perfectly well without such enquiry" (2017, p. 4246). On a more proactive turn, Newen (2017) has proposed a multi-criteria (or example-based) model that articulates a provisional mark of the cognitive by foregrounding some core methods, concepts, explanatory strategies, and frameworks that are traditionally salient in cognitive science. This strategy is put forward, argues Newen (2017, pp. 4252–4253), to avoid deep deficiencies associated with more ambitious definitional programs. Interestingly, one of the ambitious definitional programs that Newen (2017, p. 4256, p. 4266) tries to avoid is precisely that which that presupposes that any proper definition of cognition must arrive at a natural kind identification. On Newen's definitional approach, conventionalism about cognition is an open option which should be evaluated in light of the actual development of cognitive science. AT's skepticism about needing or finding a natural kind definition of cognition, and even its conventionalist stance, are, as we can see, not without company in current philosophy of cognitive science. None

of these views, despite their skepticism, predicts or advocates for the degeneration of cognitive science.

Objection 2 (O2) Conventionalism motivates some kind of pluralism about the mark of the cognitive; therefore, conventionalism is advocating for the disunity of cognitive science.

Reply Conventionalism does motivate a definitional and methodological pluralism about cognition and, given this, it can be fairly said that conventionalism is skeptical about cognitive scientific unity, if unification has to be achieved by a general consensus around a natural kind-based definition of cognition. Nevertheless, conventionalism is not "advocating" for disunity in the sense that it is causing or inducing it; disunity in cognitive science is already there! And pluralism, in a disunified scenario, might well be an advisable strategy.

The most visible symptoms of disunity are the many (and growing) fundamental, persistent disagreements about cognition at the heart of the cognitive science community. Talks of "representation wars" (Clark 2015), "border wars" or "borderline disputes" (Akagi 2017; Buckner and Fridland 2017), and debates about the "scope" of the cognitive (Allen 2017), to name but a few, are evidence of such disunity. These disagreements, we think, are not just abstract and purely speculative; they are genuine empirical struggles insomuch as they are based on, and motivate, alternative research programs. That is why these struggles have sometimes been characterized as "wars" or "crises" in the field of cognitive science (Buckner and Fridland 2017; Clark 2015).

This general state of cognitive science, its deep and persistent disputes, its multiple "crises" (Buckner and Fridland 2017), suggest, in the view of several participants and observers, that the discipline is surely approaching (if not already undergoing) a revolutionary period (Allen 2017, p. 6; Ramsey 2007; Hutto and Myin 2013). In this scenario some argue that a pluralist approach to the mark of the cognitive is not just an open option that cannot be ruled out by default, but additionally a healthy methodological recommendation to overcome programmatic dogmatism. Allen (2017, p. 4242), for example, advocates for a relaxed (but not lazy) pluralism in order to keep arbitrary conservatism at bay. In the current revolutionary state of cognitive science, argues Allen, definitive all-or-nothing marks of the cognitive are both imprudently optimistic and unnecessarily constrictive for empirical research. On the same page Akagi (2017) argues that progress in the cognitive sciences (understood as a theoretical response to new evidence) cannot avoid change to the fundamental concepts of the discipline. He believes, for this reason, that a sectarian strategy based on defining strictly dichotomous accounts of cognition is uncalled for. He instead proposes an ecumenical methodological strategy capable of mapping both widespread definitional agreements and borderlines cases of disagreement about cognition (Akagi 2017, pp. 3563–3567).

This kind of pluralism, as Allen argues (2017), is not to be identified with a lazy "anything goes" strategy. If, following AT´s approach, we understand cognition as a conventional kind concept, opening up a pluralist strategy for cognitive science, we would be equally forced to have a complementarily set of objective criteria to constantly evaluate the scientific utility of the concept, depending on the specific pragmatic and/or epistemic role we expect it to play. All in all, taking into account the

currently disunified state of cognitive science, pluralism seems to be a well-motivated strategy that deserves to be analyzed further. In this context AT's pluralism, via conventionalism about the mark of the cognitive, strikes us as an option worth considering.

References

Abramova, K., & Villalobos, M. (2015). The apparent Ur-intentionality of living beings and the game of content. *Philosophia, 43*(3), 651–668.

Akagi, M. (2017). Rethinking the problem of cognition. *Synthese, 195*(8), 3547–3570.

Allen, C. (2017). On (not) defining cognition. *Synthese, 194*(11), 4233–4249.

Barker, M., & Velasco, J. (2013). Deep conventionalism about evolutionary groups. *Philosophy of Science, 80*(5), 971–982.

Brigandt, I. (2003). Species pluralism does not imply species eliminativism. *Philosophy of Science, 70,* 1305–1316.

Buckner, C., & Fridland, E. (2017). What is cognition? Angsty monism, permissive pluralism(s), and the future of cognitive science. *Synthese, 194*(11), 4191–4195.

Clark, A. (2015). Predicting peace: The end of the representation wars—a reply to Michael Madary. In T. Metzinger & J. M. Windt (Eds.), *Open MIND: 7(R)*. Frankfurt am Main: MIND Group.

Cleland, C., & Chyba, C. (2002). Defining "life". *Origins of Life and Evolution of the Biosphere, 32*(4), 387–393.

Darwin, C. (1859/2003). The origin of species. New York: Barnes & Noble Books.

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences, 4,* 97–125.

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

Di Paolo, E., & Thompson, E. (2014). The enactive approach. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 68–78). New York: Routledge Press.

Froese, T., & Stewart, J. (2010). Life after Ashby: Ultrastability and the autopoietic foundations of biological individuality. *Cybernetics & Human Knowing, 17*(4), 83–106.

Guala, F. (2014). On the nature of social kinds. In M. Gallotti & J. Michael (Eds.), *Perspectives on social ontology and social cognition. Studies in the philosophy of sociality* (Vol. 4, pp. 57–68). Dordrecht: Springer.

Hutto, D., & Myin, E. (2013). *Radicalizing enactivism*. Cambridge: MIT Press.

Khalidi, M. A. (2013). Three kinds of social kinds. *Philosophy and Phenomenological Research, 90,* 96–112.

Machery, E. (2012). Why I stopped worrying about the definition of life… and why you should as well. *Synthese, 185,* 145–164.

Maturana, H. (1975). The organization of the living: A theory of the living organization. *International Journal of Man–machine studies, 7,* 313–332.

Maturana, H. (1981). Autopoiesis. In M. Zeleny (Ed.), *Autopoiesis: A theory of living organization* (pp. 21–33). New York; Oxford: North Holland.

Maturana, H. (2002). Autopoiesis, structural coupling and cognition: A history of these and other notions in the biology of cognition. *Cybernetics and Human Knowing, 9*(3–4), 5–34.

Maturana, H. (2003). The biological foundations of self-consciousness and the physical domain of existence. In N. Luhmann, H. Maturana, M. Namiki, V. Redder, & F. Varela (Eds.), *Beobachter: Konvergenz der Erkenntnistheorien?* (pp. 47–117). München: Wilhelm Fink Verlag.

Maturana, H. (2011). Ultrastability… autopoiesis? Reflective response to Tom Froese and John Stewart. *Cybernetics and Human Knowing, 18*(1–2), 143–152.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living*. Dordrecht, Holland: Kluwer Academic Publishers.

Mishler, B. (1999). Getting Rid of Species? In R. A. Wilson (Ed.), *Species: New Interdisciplinary Essays* (pp. 307–316). Cambridge: MIT Press.

Newen, A. (2017). What are cognitive processes? An example-based approach. *Synthese, 194*(11), 4251–4268.

O'Regan, K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*(5), 883–917.

Pöyhönen, S. (2013). Natural kinds and concept eliminativism. In V. Karakostas & D. Dieks (Eds.), *EPSA11 perspectives and foundational problems in philosophy of scien*ce (pp. 167–179). Cham: Springer International Publishing AG, (The European Philosophy of Science Association Proceedings; vol. 2).

Ramsey, W. D. (2007). *Representation reconsidered*. Cambridge: Cambridge University Press.

Rupert, R. (2013). Memory, natural kinds, and cognitive extension; or, martians don't remember, and cognitive science is not about cognition. *Review of Philosophy and Psychology, 4*(1), 25–47.

Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the Cognitive Sciences, 4,* 407–427.

Thompson, E. (2007). *Mind in life: Biology, phenomenology and the sciences of mind*. Cambridge: Harvard University Press.

Thompson, E. (2011). Replies to commentaries. *Journal of Consciousness Studies, 18*(5–6), 176–223.

Thompson, E., & Stapleton, M. (2009). Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi, 28,* 23–30.

Tirard, S., Morange, M., & Lazcano, A. (2010). The definition of life: A history of an elusive scientific endeavor. *Astrobiology, 10*(10), 1003–1009.

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

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge: MIT Press.

Villalobos, M. (2013). Enactive cognitive science: Revisionism or revolution? *Adaptive Behavior, 21*(3), 159–167.

Villalobos, M., & Dewhurst, J. (2018). Enactive autonomy in computational systems. *Synthese, 195,* 1891–1908.

Villalobos, M., & Silverman, D. (2018). Extended functionalism, radical enactivism, and the autopoietic theory of cognition: Prospects for a full revolution in cognitive science. *Phenomenology and the Cognitive Sciences, 17*(4), 719–739.

Villalobos, M., & Ward, D. (2015). Living systems: Autopoiesis, autonomy and enaction. *Philosophy and Technology, 28*(2), 225–239.

Weber, A., & Varela, F. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences, 1*(2), 97–125.