

# *Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency*

# Xabier E. Barandiaran

# Topoi

## An International Review of Philosophy

ISSN 0167-7411

Topoi  
DOI 10.1007/s11245-016-9365-4



**Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.**

# Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency

Xabier E. Barandiaran<sup>1</sup>

© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** The concept of “autonomy”, once at the core of the original enactivist proposal in *The Embodied Mind* (Varela et al. in *The embodied mind: cognitive science and human experience*. MIT Press, Cambridge, 1991), is nowadays ignored or neglected by some of the most prominent contemporary enactivists approaches. Theories of autonomy, however, come to fill a theoretical gap that sensorimotor accounts of cognition cannot ignore: they provide a naturalized account of normativity and the resources to ground the identity of a cognitive subject in its specific mode of organization. There are, however, good reasons for the contemporary neglect of autonomy as a relevant concept for enactivism. On the one hand, the concept of autonomy has too often been assimilated into autopoiesis (or basic autonomy in the molecular or biological realm) and the implications are not always clear for a dynamical sensorimotor approach to cognitive science. On the other hand, the foundational enactivist proposal displays a metaphysical tension between the concept of operational closure (autonomy), deployed as constitutive, and that of structural coupling (sensorimotor dynamics); making it hard to reconcile with the claim that experience is sensorimotorly constituted. This tension is particularly apparent when Varela et al. propose Bittorio (a 1D cellular automata) as a model of the operational closure of the nervous system as it fails to satisfy the required conditions for a sensorimotor constitution of experience. It is,

however, possible to solve these problems by re-considering autonomy at the level of sensorimotor neurodynamics. Two recent robotic simulation models are used for this task, illustrating the notion of strong sensorimotor dependency of neurodynamic patterns, and their networked intertwinement. The concept of habit is proposed as an enactivist building block for cognitive theorizing, re-conceptualizing mental life as a habit ecology, tied within an agent’s behaviour generating mechanism in coordination with its environment. Norms can be naturalized in terms of dynamic, interactively self-sustaining, coherentism. This conception of autonomous sensorimotor agency is put in contrast with those enactive approaches that reject autonomy or neglect the theoretical resources it has to offer for the project of naturalizing minds.

**Keywords** Autonomy · Enactivism · Operational closure of the nervous system · Sensorimotor constitution of experience · Mental life · Habit · Sensorimotor autonomous agency · Autonomist sensorimotor enactivism

## 1 Introduction

Enactivism is maturing and diversifying as a theoretical framework. The revolutionary narrative that has always accompanied enactivism is now warming up. Talk of “radicalism versus conservatism” or “revolution versus reform” (Chemero 2009; Hutto 2005; Hutto and Myin 2012; Thompson 2011; Thompson and Varela 2001) is growing, to the extent that some of the philosophical discussion has become almost ideological, reclaiming “authentic radicalism versus revisionism” and even fearing “negotiations with the ‘representationalist enemy’” (Vilalobos 2013, p. 163). Disputes, within what has long being

✉ Xabier E. Barandiaran  
xabier.academic@barandiaran.net

<sup>1</sup> Department of Philosophy, IAS-Research Center for Life, Mind, and Society, School of Social Work, University of the Basque Country UPV/EHU, C/Los Apaiz, 2, 01006 Vitoria-Gasteiz, Araba, Spain

a relatively small community with little internal friction, are now becoming widespread. So is the diversification of approaches. And the tensions between the internal “factions” of a formerly “unified” front start to rise. In a sense, this is all good news for it pushes the agenda forward, and this tension should be welcome. “Political” and theoretical conflict forces us to sharpen research programmes, advance more specific hypothesis, propose more accurate principles, provide more sophisticated models, etc. But it is also time to reconsider the extent to which radicalist slang and confrontational rhetoric is proving valuable or, instead, becoming an obstacle for a rigorous and systematic progress in the field. Confrontational clash is often traumatic and it risks leaving core conceptual collateral victims on its way. The concept of autonomy is one of them. Being itself one of the central claims of the original enactivist theoretical framework (Varela et al. 1991), autonomy is currently under explicit or implicit theoretical dispute and disorientation. And it is the very autonomy of enactivism that is at stake here, if it is to stand as a proper scientific paradigm within cognitive science without being dragged back into a reformed version of representationalism or subsumed under new (or old) forms of behaviourism or physicalism. The question is not trivial, and it permeates the field of philosophy of mind beyond the enactivist territory. If, as some externalists and enactivists defend, we are to reject a clear-cut agent-environment dichotomy and thus their representational relationship, how are the identity of a cognitive subject and the worldliness of its experience constituted?

The goal of this paper is twofold. The first is somewhat therapeutic: to acknowledge the conceptual tension that has historically unfolded between the concept of autonomy (both in its autopoietic form and the more abstract form of operational closure) and that of sensorimotor coupling within enactivism. The second is propositional: a re-appraisal of a notion of autonomy that can do justice to the sensorimotor constitution of cognition.

The paper is structured as follows: (a) I will review the role that the concept of autonomy has historically played within the enactivist program (and in cognitive science more generally) and how it is found absent from some recent developments, (b) I shall clarify some foundational misunderstandings or conceptual obstacles that have made autonomy a difficult notion to assume for some sensorimotor enactive approaches: (1) the overemphasis on the life-mind continuity thesis (under the “cognition = life” slogan), (2) the obscure (and yet crucial) relationship between operational closure and structural coupling, and (3) the lack of good models of sensorimotor autonomy; (c) next, I introduce autonomy back at the roots of enactivism by advancing first a notion of sensorimotor constitution of neurodynamic patterns, then translating it,

through the notion of habit, into a theoretical alternative building block for cognitive theorizing, to finally introduce the notion of sensorimotor autonomous agency; to close the paper, (d) I discuss other enactivist proposals under the light of a “autonomist sensorimotor enactivism”.

## 2 Enactivism and Autonomy

Enactivism (Stewart et al. 2010; Varela et al. 1991) has built itself mostly in opposition to the dominant computational-representationalist paradigm. The rejection of the linear sequence “Sense (input) → Plan (compute) → Action (output)” where cognition stands on the “planing” side, often under the form of symbolic propositional (or otherwise content driven) computational processing, is a unifying theme for enactivism. Together with the rejection of strong representationalism comes the emphasis on the dynamics of agent-environment sensorimotor coupling, the primacy of embodied interaction as *constitutive* of cognition.<sup>1</sup> From the most basic forms of behaviour to the highest end of human consciousness, enactivist principles provide a research agenda that displaces the focus of explanatory attention from internal representational content management to direct sensorimotor interaction. It is both in the degree of rejection of representationalism and the mode of articulating the alternative proposal (beyond the generic claims) that enactivist positions start to differ. But embodiment and sensorimotor coupling are not the only key players on the enactivist field.

A first approximation based on a dictionary definition reveals that the term “to enact” means both “to act from within” and also “to establish by law”. The very term calls for “autonomy” (from the Greek *auto* = self and *nomos* = law, norm) as a concept that integrates both the emergence or constitution of a subject, an agentic identity

<sup>1</sup> To put it in negative terms, the statement that “cognition is constituted by embodied interactions” means that cognition is not something we can hope to understand by studying brains, logical operations or information processing without continuously bringing the body and its interaction with the environment into the explanation. However, the exact extension of what constitutes minds (how much of the body, how far reaching the interaction into the environment) and what exactly is meant by constitution is often imprecise. The term constitutive has been widely used in enactivist approaches (from Maturana and Varela to Alva Noë) but a clear definition of the term is rarely provided in the enactive literature. For the purpose of this paper, by constitutive, we mean that a process or structural component is an essential part of the category *A* of objects *a* under definition, where essential implies that no other type of processes/component could do for an entity to be a member of the category *A*. Another way to put it is that *x* is constitutive of *a* iff *A* supervenes on *X* type processes/components (plus other *Y*, *Z*... type of processes) and entity *a* would not be a member of the *A* category without *x*.

that is the locus of action-perception, and the establishment (by this very subject) of its own norms of operation. These should not be understood only in the sense of bare nomological regularities but also in the proper sense of natural normativity: as the capacity to determine what is good or bad, adaptive or maladaptive, right or wrong, etc., and to regulate itself accordingly (more on this latter). *The Embodied Mind* (Varela et al. 1991, TEM hereafter), often considered the foundational text for enactivism, was particularly sensitive to the notion of autonomy. Varela et al. (1991, p. 139) stress “the necessity of understanding cognitive systems not on the basis of their input and output relationships but by their *operational closure* (...) a way of specifying classes of processes that, in their very operation, turn back upon themselves to form *autonomous networks*” (italics added). Explicit reference to Varela’s *Principles of Biological Autonomy* (1979) was repeatedly made in TEM for a more detailed account of this notion of operational or organizational closure:

We shall say that autonomous systems are organizationally closed. That is, their organization is characterized by processes such that (1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist. (Varela 1979, p. 55).

Autonomy, thus, refers here to a type of process organization that is constituted as a network of interdependent processes, where the behaviour of the whole emerges from the interaction dynamics of its component parts in a self-organized (i.e. non-instructive, symbolically structured or otherwise functionally pre-specified) manner. Autonomous systems are capable of maintaining their very organization homeostatic (that is, keeping a higher order organizational stability in the face of continuous lower order variations and perturbations). In other words, autonomous systems are recursively self-sustaining networks.

The most basic type of autonomy (autopoiesis) is found in the minimal form of life: the cell. Cellular metabolism provides a concrete, material and foundational example of a network of processes (chemical reactions, molecular self-assembling, etc.) that produce and sustain each other, creating, in turn, a boundary or membrane that encapsulates the system as a unity in the chemical domain (Ruiz-Mirazo and Moreno 2004). By doing so every cell co-determines or co-specifies, out of an in-principle undifferentiated surrounding, the set of chemical components, physical parameter ranges (e.g. temperature and pressure), or types of perturbations that constitute its “relevant” environment. Which chemical component is poisonous or how much pressure or temperature is good or bad for a

given cell, how the cell reacts to (or ignores) specific surrounding conditions, etc. is something that only the whole molecular-metabolic organization of the cell can teach us. Thus understood, autonomy becomes essential to “move away from the idea of the world as independent and extrinsic to the idea of a world as inseparable from the structure of these processes of self-modification.” (Varela et al. 1991, p. 139). This co-determination of the world of an agent by its specific mode of becoming autonomous is central to the original formulation of the enactive approach. From an autonomist perspective cognition arises when an autonomous system couples to the environment: “cognition in its most encompassing sense consists in the enactment or a bringing forth of a world by a viable history of *structural coupling*” (Varela et al. 1991, p. 205 italics added).

Moreover, autonomy, as a characterization of the fundamental logic of the living, also provides for the anchoring of the *embodiment* of cognition, in a double sense: “the body as a lived, experiential structure and the body as the context or milieu of cognitive mechanisms” (Varela et al. 1991, p. 16). Varela et al. also argued for higher levels of autonomy to be relevant for cognition, in particular, the autonomy of the nervous system: a self-organized network of recurrent dynamics mediating between sensory and motor surfaces (more on this latter) embedded, in turn, on its biological autonomous body.

As we shall see the role of autonomy for enactivism is not without its problems, but its centrality in TEM is undeniable. And yet, despite the centrality of autonomy for the early and foundational enactivism, some of the most prominent contemporary approaches—particularly the so called “radical enactivism” (Hutto and Myin 2012) and the sensorimotor contingency theory (Noë 2004; O’Regan and Noë 2001)—have either forgotten or neglected the notion of autonomy. Within these accounts, both the nature of the self to which cognitive properties are predicated, or the nature of the norms that guide or frame its behaviour are often presupposed or, simply, remain unquestioned. And this is what theories of autonomy, in a variety of flavours (Barandiaran 2004; Barandiaran and Ruiz-Mirazo 2008; Bechtel 2007a; Christensen and Hooker 2000; Di Paolo 2004; Moreno et al. 2008; Ruiz-Mirazo and Moreno 2004, 2011; Smithers 1997; Varela 1979) originally came to contribute: a theory for the emergence of identity or individuality and the norms or principles that act as sources of value or guidance.

Regarding the issue of identity, theories of autonomy have addressed the following type of questions: How does a *locus of perception-action* emerge out of a dynamical continuum or an undifferentiated mechanical ensemble? How is the *subject* of experience rooted in nature? How is a *point-of-view* fixed in the realm of natural laws? Regarding the issue of norms the problems addressed revolve around

the definition of the *adaptive* boundaries of the agents, the specification of *proper* functions in an organism, the *adequacy* of its behaviour, the *mistakes* it tries to avoid, the *evaluative framework* it is embedded in so as to drive its learning processes, the *principles* that operate on expanding or reducing potential courses of action, etc. It is important to note that, although these two questions of individuality and normativity can be treated separately (in particular, Maturana's approach, as a forerunner of enactivism, rejects normativity—see Villalobos 2013) theories of autonomy tend to bring them together and propose models of autonomous systems as candidate *organizational specifications of how individuals and norms emerge out of a network of processes that don't (necessarily) imply, assume or posses such properties*.

The sense of normativity that underlies autonomy is not to be confused with nomological processes that are applicable to raw physical processes, such as planetary motion or electromagnetic waves, whose behaviour can be fully subsumed under a law. Instead, normativity refers to the goals or principles of systemic regulation that determine good or bad functioning, the adequate or inadequate, the adaptive or the maladaptive, establishing a reference state or condition that the system strives to approach or sustain but can nevertheless fail to achieve (for a detailed treatment of the normative nature of behaviour see Burge 2009; Frankfurt 1978). The naturalizing strategy that autonomists follow can be interpreted as an organizational version of etiological approaches to normative (or proper) function and teleology; of which the historical/evolutionary line of theoretical development has attracted most attention (Millikan 1989; Wright 1976). Whereas evolutionary etiological, also called selected-effect, approaches define the proper function of a trait in terms of the effect by which that trait was selected-for along the evolutionary history of the organism (i.e. the function that made that trait an adaptation), autonomists naturalize norms within the context of the very working of the organism. The normative function of a trait or part of an organism is defined by the specific way in which it contributes to the self-maintenance of the system (Barandiaran and Egbert 2013; Barandiaran and Moreno 2008; Christensen and Bickhard 2002; Di Paolo 2005; Weber and Varela 2002). The normative nature is justified as a condition of possibility: were structure S not functioning according to the norm presupposed by the organization of the system, both the organism O (and the structure S it upholds) would not persist.<sup>2</sup>

<sup>2</sup> So, for instance, the normative (or proper) function of the heart is to pump blood at a certain rate  $X \pm n$ , because the rest of the organism relies on this pumping frequency to persist. If the heart pumps at a different rate, say  $Y$  (where  $Y < X - n$ ) the organism will die. The heart can function at different rates but  $X \pm n$  defines the viability limits or constraints of its normative functioning, above or below

Different levels or types of autonomy make it possible to specify different types of norms. So, for instance, biological autonomy (the capacity of an organism to recursively maintain its own biochemical organization: tissues, organs and their coordinated functioning) defines the level of biological norms, whereas a specific level of sensorimotor-neurodynamic autonomy would specify cognitive or behavioural normativity (how to make sense of this level of autonomy is explained in Sect. 4 in this paper).

The specification of how a holistic network of interdependent processes recursively (re)produces the conditions for its own continuation, while distinguishing itself and co-defining its environment (by specifying the types of encounters or interactions that are possible and relevant for its own continuation), makes possible the depiction of the emergence of what autonomist enactivists often call a *phenomenological domain*.<sup>3</sup> Depending on whether one opts for a monist or a pluralist approach to autonomy there can be one or various phenomenological domains, norms and identities. *Monism* accepts only a single type of autonomy (generally the physical or autopoietic) but conceives different forms (or species) of achieving such autonomy, including second or third order organizations, understood as ensembles of autonomous systems (e.g. multicellular organisms) or ensembles of ensembles of autonomous systems (e.g. an ant colony). For monists such higher order ensembles appear always subordinated to the one and only form of autonomy that ultimately anchors the normative and regulatory principles of the higher orders (see Christensen and Hooker 2000 for a prototypical example). A *pluralist* approach opens the way to conceive of different types of autonomy: the autonomy of the cell (Bechtel 2007a; Ruiz-Mirazo and Moreno 2004), the autonomy of multicellular organisms (Arnellos et al. 2014; Rosslenbroich 2009), the autonomy of behaviour (Barandiaran 2004; Barandiaran and Moreno 2006b; Smithers 1997), the autonomy of inter-subjective interaction (De Jaegher and Di Paolo 2007; De Jaegher et al. 2010) the autonomy of the social (Luhmann 1986, 1995), or the political (Adams 2007; Castoriadis 1997), etc. At each level a new phenomenological domain is opened, new forms of identity and novel types of norms emerge. Thus, theories of autonomy provide not only with the specification of a domain of interactions and a

Footnote 2 continued

those limits the heart will malfunction, it will operate against the norm.

<sup>3</sup> Autonomist enactivists consider that a phenomenological domain is opened whenever there is an autonomous (or operationally closed) unit. Since, out of an undifferentiated environment, the specific mode of closure or autonomy of a system (its self-producing or sustaining network of processes) defines the form of structural coupling with the environment (its membrane properties and metabolically relevant compounds, its sensorimotor habitat, etc.) each form of autonomy anchors a phenomenological domain, that is, a selectively relevant environment, a form of “experiencing” or encountering it.

naturalized source of identity and norms but for a reference framework from which relational or holistic categories can be naturalized: function, value, regulation, meaning, intentionality, adaptation, will, and others.

Whether we assume a monist or a pluralist conception, the core explanatory role that autonomy is called on to play in cognitive science is apparent. And yet, recent emphasis on the sensorimotor nature of the enactive mind has left autonomy out of focus, centring the debate around a definitive victory over the notion of representation (Hutto 2005; Hutto and Myin 2012; Menary 2006), the clarification of the sensorimotor constitution of experience and the nature of skills and knowledge of sensorimotor contingencies (Noë 2004; O'Regan and Noë 2001) or, more broadly, the conception of cognition as "subserving action and being grounded in sensorimotor coupling" (Engel 2010; Engel et al. 2013) assuming that "ultimately, there is no prospect of understanding minds without reference to interaction between organisms and their environments" (Hutto and Myin 2012, p. 4). These approaches conceive of themselves as enactivists, explicitly following TEM's call to study cognition in terms of embodied action, and reclaiming the *canonical definition of enactivism*:

The enactive approach consists of two points: (1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided. (Varela et al. 1991, p. 173).

The focus of explanatory resources here is the primacy of action for cognition, the way in which motor variations induce (via the environment) sensory variations and the kind of regularities that emerge out of recurrent sensorimotor loops as constituting experience or organizing cognitive life. The strongest claim of enactivism is the *sensorimotor constitution of experience*: "perceiving is *constituted* by the exercise of a range of sensorimotor skills" (Noë 2004, p. 90 *italics added*). The interpretations of this thesis range from the most radical and direct, in which perceptual experience—or the experience of space itself (Lenay and Steiner 2010)—depends directly on sensorimotor dynamics, on actions regularly and recurrently transforming the sensory stream through the environment, up to the most abstract in which the involvement of skills or knowledge of sensorimotor dependencies might suffice. These forms of enactivism are often intensely concerned with the rejection of internalism (understood as the thesis that cognition is constituted or explained solely with respect to structures and properties internal to the organism).<sup>4</sup> It is the structure of sensorimotor regularities

that body-environment coupling affords, rather than the operational closure of living organization, that sets up the research agenda. As a result, it has turned out to be conceivable, even popular, to defend enactivism without reference to autonomy. The pressing question here is *why* autonomy was lost somewhere along the way towards some of the contemporary positions, when the very idea was precisely situated at the roots of the original proposal. Answering this question should help us see whether to reject autonomy as an explanatory resource for contemporary enactivism or to enrich the research program with an updated version.

In order to clarify the theoretical landscape ahead I will use the label "autonomist enactivism" to cover those approaches that take bodily (physiological-metabolic) and neuronal self-organizing properties as critical for enactivist theorizing and I will use "sensorimotor enactivism" (or simply enactivism) to cover those that centre their approach on sensorimotor coupling, contingencies or coordinations.<sup>5</sup> Note that these varieties of enactivism compose intersecting but non-overlapping sets. One can be a sensorimotor enactivist but pay no attention to autonomy. It is also possible to root enactivism in autonomy without considering a proper level of sensorimotor autonomy. The view defended here is precisely the centrality of the intersecting domain, that is, the space of "autonomist sensorimotor enactivism" where autonomy is depicted at the very level of sensorimotor dynamics.

#### Footnote 4 continued

sensory variation depends on motor movement), or the way in which perception is action dependent, what drives the attention of sensorimotor enactivism is the structure of the environment as enacted by a particular type of embodiment (Block 2005; Lenay and Steiner 2010; Noë 2004; Prinz 2006; Rowlands 2009a). This has put enactivism in direct confrontation with internalists who believe that the nature or the mark of the mental is the capacity of the agents to internally represent the external environment (Adams and Aizawa 2009; Aizawa 2007; Rowlands 2009b).

<sup>5</sup> Other labels to classify different varieties of enactivism are taking momentum. Some of the most popular include "radical enactivism" and "autopoietic enactivism". The first has been used to refer to those forms of sensorimotor enactivism whose explanatory strategies do not involve any reference to representational or informational content (Hutto 2005, Hutto and Myin 2012). This "negative" definition is, I suggest, but one way of being radical and the way in which the label has been used so far falls short of acknowledging and providing a positive radical research program (more on this in the discussion section). "Autopoietic enactivism" is the most widespread label to characterize Maturana and Varela's approach and its latter developments (Thompson 2007, Stewart et al. 2010). Since autopoiesis is the name that Maturana and Varela coined for the most basic or material (physico-chemical or metabolic) form of autonomy (or operational closure) this label is often confusing for it masks that there are other levels of autonomy and, particularly, it often ignores neurodynamic and sensorimotor autonomy.

<sup>4</sup> By putting the explanatory burden on the way in which an agent executes its mastery of sensorimotor contingencies (the way in which

### 3 Autonomy: What Went Wrong

It is possible to identify three main obstacles that justify the progressive distance between theories of autonomy and some of the contemporary enactivist “schools”. First, the overemphasis on a narrow conception of autonomy, metabolic or autopoietic autonomy, as a fall-back position to ground cognitive phenomena. Second, the notion of “operational closure of the nervous systems”, as expressed in the early writings, turns out to be difficult to reconcile with the nowadays more fashionable claim that “minds ain’t in the head” and the emphasis on the sensorimotor constitution of experience (Noë 2004). And, finally, the lack of a good model of sensorimotor autonomy. In contrast with the operationally explicit models of basic or material autonomy that have provided maturity and empirical support for autopoietic theory at the level of metabolic or molecular interactions (Barandiaran and Ruiz-Mirazo 2008; Bechtel 2007a; Bourgine and Stewart 2004; Luisi 2003; Piedrafita et al. 2010; Varela et al. 1974), early models of neuronal or sensorimotor autonomy (in particular Bittorio as a model of the operational closure of the nervous system, to be discussed below) failed to provide conceptual consistency within enactivism. By analysing these difficulties in detail we might gain not just a therapeutic grasp on how to reconcile autonomy back with enactivism but we might also be capable to specify the demands for a more precise characterization of autonomy that becomes useful and compatible with contemporary enactivist approaches.

#### 3.1 The Problem of “Cognition = Life” Thesis

One of the strongest positions within the enactivist family of views, often associated with an autonomous approach to enactivism, is the claim that “Cognition = Life”. It can be found in Maturana’s early claim that “[l]iving systems are cognitive systems, and living as a process is a process of cognition. This statement is valid for all organisms, with and without a nervous system” (Maturana 1970, p. 13). In the foundational document of the autopoietic theory we also find a similar statement: “The domain of all the interactions in which an autopoietic system [a living system] can enter without loss of identity is its cognitive domain” (Maturana and Varela 1980, p. 119). This conception of cognition reached *The Embodied Mind* and it has since permeated what Hutto and Myin label as “autopoietic enactivism” (Hutto and Myin 2012) that has many contemporary variations (Bourgine and Stewart 2004; Di Paolo and Thompson 2014; Froese and Di Paolo 2011; Thompson 2007).

The life-cognition identity thesis, as we might label it, takes autopoietic autonomy (or any level of autonomy) and

makes of it a sufficient condition for the predication of cognitive capacities: sense-making, intentionality, worldliness, phenomenological experience, etc. This loose characterization of cognition has lessened the attractiveness of “autonomy” for some enactivists. There are good reasons for this. A strict interpretation of the claim “life = cognition” makes epileptic attacks, or human vegetative life, breathing, digestion or falling down a cliff (right until the bottom is reached) all members of the cognitive club. As far as the system remains alive they are all instances of “structural coupling of a living system without loss of autopoiesis”. To avoid this consequence, certain amendments to the club membership have been made along autopoietic enactivism. A more charitable interpretation demands that “Cognition = autopoiesis + adaptive interaction” (Bourgine and Stewart 2004), so that not just any behaviour (or structural coupling) that “doesn’t kill you” would count as cognitive, but one that actually improves your chances for self preservation, so that your action is guided towards an adaptive goal or away from a “dangerous” situation. However, this still leaves breathing (no matter how automatic), plant roots growing towards a more humid ground, or body temperature regulation, as genuinely cognitive phenomena.

Moreover, it is not always clear what the role of life really is for cognition when the researcher comes to explain a specific example of cognitive behaviour. In addition, it leaves the pressing question open of whether mindful behaviour is at all possible without direct biological relevance. So if, for instance, artificially activated thirst generates a genuine intention to drink, and the glass of water has no adaptive consequences on our body, suddenly, the thesis seems to imply, we are devoid of any genuine intentionality when it comes to grasping and drinking from the glass. This seems at odds with our experience. Moreover, many instances of cognition seem to lie out of the realm of the biologically (i.e. autopoietically) relevant, and yet, they are fully characterizable as cognitive: surfing waves, skilfully mastering backgammon or, in a more extreme case, learning how to kill yourself or assuming a strategic risk that, incidentally, happens to lead to your death. These are all cognitively demanding tasks that, nevertheless, bear little contribution to autopoiesis.

It seems rather intuitive that biological and cognitive norms and identities are not always coextensive. Cases of double personality within the same body (Reinders et al. 2003, 2006) are one example of the latter. At the level of norms it is arguable that cognitive failure does not necessarily imply biological failure in any sense that is directly relevant for autopoiesis. In other words, cognitive normativity remains underspecified by biological normativity and it is often hard for cognitive science to see an explanatory

gain from the life-cognition identity thesis (for a detailed development of these problems see Barandiaran 2008, Chapter 7).

If taken as a *sufficient* condition, it is not clear for some contemporary enactivist schools why and, more importantly, *how* autopoiesis or biological-metabolic autonomy is relevant to accounting for sensorimotor contingencies or sensorimotor coupling more generally. On the other hand, if taken as a *necessary* condition, then, it is not clear whether this condition is necessary *de facto* (natural cognitive systems need to be alive, if dead they are not mentally active and if non-living, e.g. a mountain or a cloud, they are not cognitive systems) or *de jure* (a robot needs to be alive for it to display genuine cognitive capacities).<sup>6</sup> But in either case there exists a relatively comfortable position for sensorimotor enactivists to displace the autonomy of the living from their explanatory tool-kit. Sensorimotor enactivists can reason as follows: “Sure, life provides a set of necessary conditions for a system to be cognitive, but everything that is explanatorily relevant about cognition is something we can explain in great detail without specific reference to autopoiesis. I need not deal with the details of metabolism in order to do cognitive science. It is perception-action loops that matter, the structure of sensorimotor contingencies, the way in which sensorimotor skills are acquired, and so on. I can certainly assume that the underlying infrastructure needs to be alive, but my research can carry on unchanged right after I accept it”.

The more general problem of the life-mind identity thesis concerns the specificity of cognition. For some autonomous monists like Christensen and Hooker (2000) or Bickhard (2000, 2009), there is only biological autonomy, and that is the ultimate origin of norms and self. What sets cognition apart from mere biology, the specificity of the cognitive, the mark of the mental, is given by the complexity of the interactions involved; and by their representational nature. In this sense what autonomy provides here is a way to naturalize epistemic norms and the emergence of a biological self whose way of self-maintenance depends on its representational capacity. This is not simply a lower level aspect that bears no consequence for cognitive capacities, on the contrary, cognition, according to these authors, is deeply rooted in biological autonomy as a source of error correction that bootstraps cognitive development. But they fail to account for the non co-extensiveness of biological and cognitive normativity and identity. In addition, they fall into a form of representationalism, close to the action-oriented category (with the additional requirement that conditions of satisfaction are tested against its consequences on biological

autonomy), that has been intensively argued against in the enactivist literature (Hutto and Myin 2012). Other monists simply adhere to the cognition-life identity thesis we just discussed above.

On the pluralist side positions appear in two flavours. For some, as we have seen, any level of autonomy affords a cognitive characterization and the specificity of cognition is thus lost: cognition (unicellular, animal, collective, socio-technical) is coextensive with the agent-environment coupling no matter the instantiation or the level of realization of the autonomous organization. There is, however, another alternative position: the specificity of cognition lies at a particular level of biological organization, one that makes possible the emergence of new form of autonomy (embedded but distinct from generic biological autopoiesis): that of the nervous system. In fact, although ambiguous at times, for Varela et al. it is the autonomy of the nervous system that provides the level of specificity that is relevant for enactivism. In addition to autopoiesis (which is never directly mentioned in TEM) what deserves special attention from the enactivist point of view is the operational closure (i.e. the autonomy) of the nervous system. And this brings us to the next set of obstacles that autonomy poses for some contemporary enactivist approaches: the formulation of the autonomy of the nervous system that Varela et al. provide in TEM (and particularly in Varela 1979) is problematic, as we are about to see, and the model they propose to illustrate the autonomy of the nervous systems turns out to fail to address the central tenet of enactivism. So, although an autonomist perspective on enactivism need not collapse into a life-mind identity thesis (as autopoiesis often seems to imply), it is not always clear how the alternative pluralist notion of autonomy is to be articulated for it to be of direct use for the enactivism programme.

### 3.2 The Problem of Non-intersecting Domains: Relational Structural Coupling and Constitutive Operational Closure

For Maturana and Varela the operational closure of the nervous system plays a prominent role since their early theorizing<sup>7</sup>:

Operationally, the nervous system is a closed network of interacting neurons such that a change of activity in a neuron always leads to a change of activity in other neurons, either directly through synaptic action,

<sup>6</sup> And, if so, it is not clear what metabolism really adds to the way in which the robot effectively operates at the level of sensorimotor dynamics.

<sup>7</sup> In fact, the very notion of operational closure, before turning into a multi-scale theory of autopoiesis and second order forms of closure with Francisco Varela, was originally developed by Humberto Maturana as applied to the operations of the nervous system (Maturana 1970).

or indirectly thorough the participation of some physical or chemical intervening element. Therefore, the organization of the nervous system as a finite neuronal network is defined by relations of closeness in the neuronal interactions generated in the network. (Maturana and Varela 1980, p. 127)

This account of closure or autonomy at the level of neuronal activity risks a straightforward solipsist interpretation that clashes with the *constitutive* role attributed to sensorimotor dynamics, externalism and action in contemporary enactivism. If the operations of the nervous system are to be understood as self-referential and closed, where any activity is always internal to the network, how is this closure to be reconciled with enactivism's central claim that cognition consists in the "recurrent sensorimotor patterns that enable action to be perceptually guided"?

The tension is evidently seen in the treatment of the role (or rather lack of it) that Varela attributes to the environment:

That an [...] observer should see environmental elements intervening between the effector and the sensory surfaces of the organism is *irrelevant*, because the nervous system can be defined as a network of neuronal interactions in terms of the interactions of its component neurons, *regardless of intervening elements*. Therefore, as long as the neuronal network closes on itself, its phenomenology is the phenomenology of a closed system in which neuronal activity always leads to neuronal activity. (Varela 1979: 242, italics added)

He reinforces this claim by stressing that some cognitive operations are independent of any feature of the environment:

"We must recognize that this effect [size constancy] corresponds to a process that takes place completely within the nervous system, *independently* of any feature of the environment, although it may be *elicited* by interactions of the organisms in its environment." (Varela 1979: 254, italics added)

On the other hand, the notion of structural coupling is made to play an opposing (at times contradictory) role. So, for instance, the state-space of the nervous system is said to be a function of its history of interactions:

"(...) the domain of the possible states that the nervous system can adopt as a closed system is at any moment *a function of this history of interactions*, and implies it." (Varela 1979: 245)

Or environmental perturbations are said to be "historical determinants" of the nervous system:

"The changes that the nervous system's structure can undergo without disintegration (...) are fully specified by its connectivity, and the perturbing agent only constitutes a historical *determinant* for the concurrence of that changes." (Varela 1979: 242, italics added)

The result seems to be an irreconcilable marriage. Either the NS-environment coupling does dynamically integrate environmental or sensorimotor regularities into the activity of the NS (whether directly or in terms of sensorimotor coordination patterns) or it does not. Either environmental features are "irrelevant", the activity of the nervous system is "*independent* of any feature of the environment" and the nervous system functions "regardless of intervening elements" or interactions are "determinants" for changes on the NS, these changes are "elicited by interactions" and the domain of states of the NS is "a function of the history of interactions". You simply can't have it both ways.<sup>8</sup>

But the problem goes deeper. The epistemological framework that underlies this conundrum perfectly illustrates the tension with enactivism. Maturana and Varela (1980) distinguish two "non-intersecting" domains: one is termed *relational* (or functional) and the other *operational* (or constitutive). The *operational domain* is conceived as mechanistic (often in terms of deterministic dynamical systems, inherited from Ashby's cybernetic legacy). This is the domain of operations between parts of a system, the domain of causal chains and mechanistic structures. When these causal chains close on themselves they constitute operationally closed (i.e. autonomous) systems. Changes in this closed network can be triggered by perturbations from its surroundings. But the fact that these changes originate in the environment, that a given change in an operationally closed system had a tree or a stone as a cause, is alien to the system that remains, in a sense, "blindly" structurally coupled with its environment. This brings us to the second domain. The *relational domain* pertains to the observer, who can access relations (e.g. correlations) between different elements, between the constituents (e.g. molecular reactions) of a system and its environment (e.g. a sugar, or food, gradient in a petri dish for a bacterium). The structural coupling between a system and its environment is also accessible as such to an observer, and thus pertains to the relational domain. Conflating relational properties (which are observer dependent) with operational properties (that belong to the constitutive domain) involves a category mistake. Conflating both domains is often, Maturana and Varela argue, the source of misconceptions in cognitive

<sup>8</sup> More generous interpretations of Varela's work can be made, of course. The point here is not whether Varela was right or wrong, but to highlight the points of highest theoretical contrast, so as to identify the boundaries of a conceptual problem that demands attention.

science and philosophy of mind, representationalism being the most prominent of such misconceptions for it takes informational content, a relational term, to be causally relevant to explaining the operational workings of an organism.

The crux of the problems is that the central thesis of some enactivists (e.g. Noë 2004) lies precisely at the crossroads of the intersection between these domains: it claims that structural coupling (placed in the “relational” domain by Maturana and Varela) is constitutive of cognition. And this is a violation of the non-intersecting axiom (see Fig. 1). Advocates of autonomist enactivism seem to be forced into an impossible choice: either they opt for autonomy without enactivism or they opt for enactivism without autonomy.

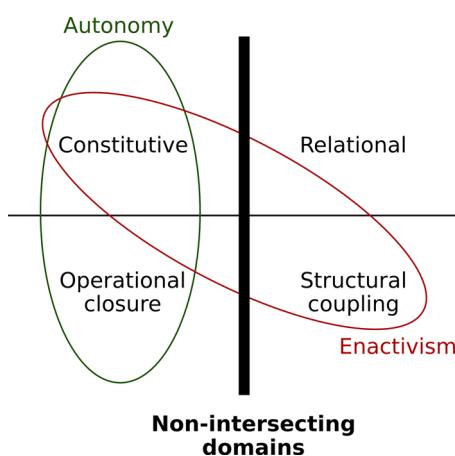
In TEM the tension between the primacy of sensorimotor coupling and the operational closure of the nervous system gave rise to a form of unstable equilibrium that oscillates between the emphasis on either one of the two. When Varela et al. discuss “hands-on applications” of enactivism in cognitive science (Varela et al. 1991, pp. 207–211), they focus on Rodney Brooks’ subsumption architecture for robots, but here the notion of operational closure or autopoiesis does not function as a design principle or play an explanatory role; raw sensorimotor coupling does all the work. On the other hand, when they provide examples of autonomy (metabolic or neural) they skip any reference to the role that sensorimotor (or structural) coupling plays in cognitive performance. Thus, for instance, Varela et al. consider “color as a form of experience that is

constituted through emergent patterns of neuronal activity” (Varela et al. 1991, p. 166)<sup>9</sup> but they make no reference to the role of sensorimotor dynamics (see Bompas and O'Regan 2006, for a sensorimotor theory of color vision).

### 3.3 The Problem of Bittorio as a Conceptual Model for Cognition

The tension between sensorimotor enactivism and autonomy is particularly apparent in TEM when the simulation model named Bittorio is introduced to illustrate the relationship between operational closure and structural coupling (Varela et al. 1991, pp. 151–157) and it perfectly illustrates the conceptual problem we have dealt with in the previous subsection. It also brings us to the last of the historical problems that might have pushed some enactivist factions away from autonomy: the lack of a good model of sensorimotor autonomy—in contrast with the operationally explicit models of basic or material autonomy (Barandiaran and Ruiz-Mirazo 2008; Varela et al. 1974), a vacuum that has reinforced the reversion to autopoiesis (or basic-metabolic autonomy) as the prominent anchoring of autonomy within enactivism.

Technically speaking, Bittorio is a one dimensional toroidal cellular automaton (Wolfram 1984) where the state transition function for each cell takes as input the state of other cells, and computes the next state of the cells synchronously. The way to illustrate a cellular automaton (CA) of this kind is a string of cells (or boxes) that can take binary states (0 or 1). Although pictured in a straight line, the state of each cell in a 1D CA can affect and be affected by the state of other cells in the string (beyond those that are adjacent to it). In addition the extreme ends of the chain are connected to each other, making a ring, so that there is no beginning or end of the CA. The state of the whole network is updated simultaneously at discrete time steps, so the next state of each cell at  $t = 1$  is computed according to the states at  $t = 0$  of all the cells it is connected to. Each transition rule is a boolean function of the kind “if the state of the cells with incoming connections is 0 and your current state is 1, your next state will be 0”. In this sense, the CA can be seen as a closed boolean interaction network where the state of each cell can be affected by and, in turn, can affect the state of any other cell (either directly or indirectly by propagating changes). The result is a simple yet complex deterministic non-linear discrete dynamical system. Small changes in initial conditions or small perturbations to the network can fully divert the course of the



**Fig. 1** The relational and constitutive domains in Maturana and Varela's framework. The autopoiesis or operational closure (autonomy) belongs to the constitutive domain, whereas structural coupling belongs to the relational domain. Both domains are considered non-intersecting and explanations that bring about components of both domains are considered to fall into a category mistake. The enactivist claim that takes sensorimotor coupling to be constitutive of cognition lies precisely at the intersection between these two domains

<sup>9</sup> Often, in the autonomist enactive literature, constitutive and interactive aspects of autonomy have been distinguished (Barandiaran, 2004; Froese et al. 2012), but this doesn't really solve the problem, for the challenge here is to conceive the behavioural or sensorimotor as constitutive.

dynamics of the network. Conversely, big differences or changes in network states can converge to the same final state or attractor.

Varela et al. embedded Bittorio in an environment composed of fixed binary states that could affect, or perturb, the state of some “sensory” cells of the network. Depending on the environmental states encountered, the transition rules that constitute Bittorio, and the current state of the network, these perturbations will affect the network differently. Varela et al. configured Bittorio so as to be capable of co-varying with or discriminating odd sequences of perturbations, not by explicitly coding a representational rule for this, but simply by tuning the transition rules so as to have the effect of “driving” the evolution of the network differently when encountering odd or even sequences of perturbations. “[O]n the basis of its autonomy (closure)” Varela et al. conclude, “[Bittorio] performs an interpretation in the sense that it selects or brings forth a domain of significance out of the background of its *random milieu*” (Varela et al. 1991, p. 156 *italics added*).

There are two important features of this model that made it particularly inappropriate for putting autonomy at the core of enactivism. First, it conceives of the environment as a “random milieu” and second, and most importantly, it provides no means by which Bittorio can affect the environment. Moreover the environment is seen just as a “source of perturbation”. As a result, Bittorio has no way of enacting sensorimotor regularities of any kind: it has no motor capacity to influence its sensory changes through the environment and the environment has no regularities to be exploited as sensorimotor contingency structures. So, although Bittorio might satisfy some of the properties attributed to autonomous systems it fails to satisfy the enactivist core statement that “cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided” (Varela et al. 1991, p. 173).

In the light of this model, and the ambiguity inherent to the relationship between the notion of “operational closure of the nervous system” and that of “structural coupling” with a “random milieu”, it appears that it is no detriment for some enactivists to ignore the notion of autonomy, but almost a forced move if theoretical consistency is to be preserved: if sensorimotor regularities need be considered constitutive of experience (to the extent of externalism), how could it be reconciled with the operational closure of the nervous system?

## 4 The Autonomy of Mental Life

The problems just outlined do not preclude a fundamental role for autonomy in enactivism. They do, however, explain the conceptual difficulties and problems that the

canonical version of autonomy might pose for contemporary forms of sensorimotor enactivism. By doing so these problems specify the job description that a theory of autonomy needs to satisfy to be compatible with sensorimotor approaches to cognition: (a) it needs to distinguish itself from raw metabolic or biological autonomy (auto-poiesis) and provide an explanatory relevance to the constitution of identity and norms at the level of sensorimotor dynamics, (b) it must be capable of specifying how cognition is constitutively sensorimotor, and (c) it must deliver models that can illustrate the concept.

For exegetical completeness it is worth noting that, in a much less well known text, Francisco Varela (1992, see also 1997) provided a revised conception of the operational closure of the nervous system that rightly opens the path for a notion of autonomy that is constitutively sensorimotor and perfectly compatible with enactivism:

I speak of “closure” to highlight the self-referential quality of the interneuron network and of the perceptuo-motor surfaces whose correlations it subserves. The qualification “operational” emphasizes that closure is used in its mathematical sense of recursivity, and *not* in the sense of closedness or isolation from interaction, which would be, of course, nonsense. More specifically, the nervous system is organized by the operational closure of a network of reciprocally related modular sub-networks giving rise to ensembles of coherent activity such that: (i) they continuously mediate invariant patterns of sensory-motor correlation of the sensory and effector surfaces; (ii) give rise to a behavior for the total organism as a mobile unit in space. The operational closure of the nervous system then brings forth a specific *mode* of coherence, which is embedded in the organism. This coherence is a *cognitive self*: a unit of perception/motion in space, sensory-motor invariances mediated through the interneuron network. (Varela 1992, p. 10)

Unfortunately, Varela did not elaborate this renewed notion of closure in more detail,<sup>10</sup> he provided no models of this advanced form of autonomy and his latter experimental work focused on perceptual consciousness where he

<sup>10</sup> A detailed unfolding of this characterization of cognitive closure that sustains behavioural autonomy is much needed, for the passage leaves far to many open questions: what is exactly the relationship of mediation between closure at the level of “modular sub-networks” and the patterns of sensorimotor correlation? Why is motility important? How is “the specific mode of coherence” constituted? Etc. Miguel Aguilera (2015) has done an excellent job, expanding upon previous work by Di Paolo (2003, 2008), Barandiaran and Moreno (2006b) and Barandiaran (2008).

studied large-scale neuronal coordination without targeting the role of sensorimotor dynamics (Varela et al. 2001).

Ten years ago I started to develop this line of research<sup>11</sup> in order to rescue a notion of autonomy that is valid and complementary to the sensorimotor nature of our mental lives (Barandiaran 2004, 2007; Barandiaran and Moreno 2006b). One way to identify this path is by highlighting that mental life is unlike other forms of life (particularly biological life) and that the analogy between life and mind need not imply a reductive (or identity) continuity thesis between the biological and the cognitive. In fact I have long advocated for a “biology  $\neq$  cognition” thesis—see (Barandiaran and Moreno 2006b) but particularly (Barandiaran 2008, Chapters 7–8). In turn, this approach distinguishes itself from the more traditional notion of operational closure of the nervous system by assuming that cognitive autonomy is constitutively sensorimotor; i.e. that the neurodynamic patterns that are characteristic of our lived experience depend on a *strong* sense of sensorimotor coupling. Mental life’s specificity lies in the fact that the environment is not simply a source of perturbations (or a source of energy and matter) for an otherwise operationally closed network. On the contrary, the nervous system’s activity achieves its closure, its large scale coherence, *through* embodied interaction: i.e. through fine grained coordinations between neurodynamic and sensorimotor correlations. I will clarify this point with a set of recent simulation models (Aguilera et al. 2013, 2015; Santos et al. 2012). These models fill the gap left open by Bittorio as a canonical model for sensorimotor autonomy and open the path for a re-appraisal of a sensorimotor enactivist sense of autonomy.

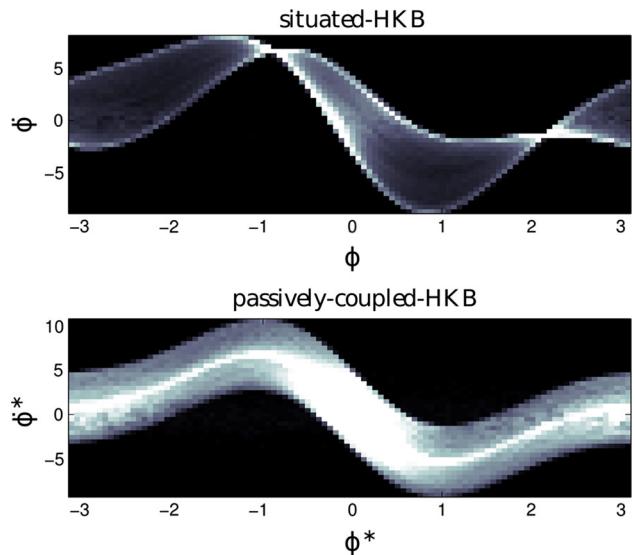
#### 4.1 Understanding the Sensorimotor Constitution of Neurodynamic Patterns Through the Situated HKB Model

In a set of recent models we have tried to show how a neurodynamic pattern could be sensorimotorily constituted. The situated-HKB model (Aguilera et al. 2013; Santos et al. 2012) shows a two wheeled robot moving towards a gradient of stimulation (what we shall call *phototactic behaviour*) in a two dimensional environment. The robot is controlled by the extended HKB equation (Kelso et al. 1990), representing the phase difference (variable  $\phi$ ) between two oscillatory components that we might interpret as sensory and motor cortices of an extremely simplified brain.

<sup>11</sup> I draw inspiration from Varela’s less cited texts, some of Di Paolo’s developments (Di Paolo 2003), together with Smithers’ intuitions on robotic autonomy (Smithers 1995, 1997) and Piaget’s sensorimotor grounding of cognitive development (Piaget 1969, 1975).

In both papers we carried out a crucial experiment that helps illustrate and clarify the notion of sensorimotor constitution of neurodynamic patterns: the sensory input of a freely behaving (sensorimotorily coupled) agent was recorded and then played back as an input into an identical robot (that we label as “partially coupled”). The resulting neurodynamic patterns were then compared. The experiment is critical to distinguish a mere causal role of the input from a constitutive role of sensorimotor coordinations. For the agent that behaves freely the sensory stimulation is the result of the effect of motor variations that in turn affect its relative position in the environment which in turn changes the sensory surface in a continuous closed loop of influence mediated by the agent’s body morphology and “brain” activity. Despite the fact that the identical twin robot receives the exact same sensory stream, both slight variations in initial conditions or small fluctuations immediately give rise to qualitatively different neurodynamic patterns for the partially-coupled agent.

As seen in Fig. 2, whereas the partially-coupled agent explores almost homogeneously the whole state space (which is exactly the thick undulating band cut against the black background), the situated agent distinctively shapes a pattern inside it. It is this shape that is functionally relevant to generate phototactic behaviour. The neurodynamic signature or neuronal correlate of phototaxis is precisely that shape, the carving out of a specific, functionally distinct, *form* within the undifferentiated space of possible oscillatory relationships. Neither the intrinsic and spontaneous



**Fig. 2** Signature of the situated HKB with  $s = 2.5$  and the corresponding passively-coupled HKB. It represents the density distribution of the effective phase space of the HKB equation when it is coupled with an environment, showing the difference between situated and passive coupling. [taken from Aguilera et al. 2013, with a CC-by license]

activity of the system, nor the shape of the system under behaviourally-structured input can produce and sustain such a pattern. It is the fine grained coordination between oscillatory brain activity and sensorimotor contingencies that sustains it. The difference between both conditions is even more apparent when comparing the robustness to sensory, motor or neurodynamic fluctuations (e.g. background noise) that the situated agent displays. Inducing noise into the situated agent does not alter its behaviour, it is able to compensate for that noise behaviourally (maintaining phototactic capacities and the corresponding neurodynamic signature). Inducing noise into the partially-coupled agent, on the contrary, provokes divergent and incoherent dynamics (for further details see Aguilera et al. 2013).

The model illustrates a case where fine-grained sensorimotor coordination is *necessary* to achieve functionally distinct neural signatures. What is a necessary condition here is *not* a specific sensory state, a perturbation or a structured or transformation-invariant sequence of them, of which the motor side is simply a consequence and, incidentally, an instrumental cause, as if other sources of sensory structuring could equally do—as Hurley has suggested is the case for sensory-dependency in ecological psychology (2001), see also Mossio and Taraborelli (2008). What is constitutive is the sensorimotor *coordination* itself. The way in which motor neurons drive, through the environment, the activity of sensors is part of the pattern-formation process (as much as the influence of other “internal” parameters or variables).

It is important to note that, in the extreme case, what matters to achieving the right neurodynamic signature is the coordination between the sensory input and the state of the order parameter  $\varphi$ , so that the sensory parameter variation modulates the intrinsic dynamic of  $\varphi$  in the “right manner”. But the only way in which this coordination can take place with fine grained temporal accuracy is precisely through sensorimotor coupling (or equivalently, by simulating the sensorimotor loop). Thus, what matters is not the sensory input, but the fine grained temporal structure of sensorimotor contingencies as they are enacted by the agent. To say it differently, even if one considers that mental states supervene on brain processes, there is a sense in which sensorimotor dynamics are constitutive. Let’s say that cognitive state Y supervenes on macroscopic brain state X, which emerges from correlations between sensory neurons S, interneurons N and motor neurons M. When I say that cognitive state Y is sensorimotorly constituted I mean that the correlation  $M \rightarrow S$  (the way in which sensory variation correlates with motor variation through sensorimotor coupling), is as much a constituent of X (and thus of Y) as the correlations  $S \rightarrow N$  or  $N \rightarrow M$ .

We can avoid committing ourselves to the externalist position that many have attributed to Alva Noë—and strongly criticized him for adopting (Block 2005; Prinz 2006)—that is, the supervenience of cognition upon (extended, out of the head) environmental states. Sensorimotor constitution involves the correlation structure of  $M \rightarrow S$  as constitutive of cognition, it needs not assume or defend that the state of the environment is also constitutive, that is  $M \rightarrow E \rightarrow S$ . It is, however, worth noting that the set of possible sensorimotor correlations that a body can enact in a particular situated condition *is* the environment *for* the agent (Buhrmann et al. 2013). It is only within this environment that certain neurodynamic patterns can be enacted, only when the sensorimotor environment is actively inhabited in this sense, when sensorimotor contingencies are enacted through sensorimotor coupling, that certain functionally distinct neurodynamic patterns emerge. They don’t, they cannot, exist (be initiated, developed, sustained and terminated) without their enactment. And this is what qualifies as sensorimotor constitution.<sup>12</sup>

Although Varela et al. might have found these models illustrative, our experiments and interpretation introduces significant improvements regarding the conceptualization of operational closure and structural coupling: (1) to depict the environment not as a source of perturbations or deformations to an otherwise operationally closed network of neuronal activity but, instead, as a necessary condition for generating and sustaining (through sensorimotor coupling) the neurodynamic patterns that are characteristic of a certain behavioural capacity, and conversely (2) to show that closure should not be predicated of a set of pre-existing states within a network, but of patterns (with their specific temporal unfolding) that can only take place when the right sensorimotor correlations are enacted. These are not outdated amendments to an old version of enactivism. Some recent autonomist enactive approaches still seem to rely on a weak conception of sensorimotor coupling<sup>13</sup>:

When a stimulus arrives, the activated receptors transmit pulses to the sensory cortex, where they induce the construction by nonlinear dynamics of an activity pattern in the form of a large-scale spatial

<sup>12</sup> The point made here is no surprise for those that have long advocated for dynamical systems theory as a formal and conceptual framework for understanding cognition. What we managed to show is a clear illustration of a simple model with just one attractor and single variable that can display a strong form of sensorimotor coupling dependency and show the kind of experiment that turns out critical to depict what exactly is meant by sensorimotor constitution.

<sup>13</sup> Thompson would most probably agree with the account of sensorimotor constitution I am defending here. This quote does not represent an opposing view but, rather, the ambiguity or lack of explicit theory of sensorimotor constitution that some authors still display, increasing the tension between autonomist enactivism and sensorimotor enactivism.

pattern of coherent oscillatory activity. This pattern is not a representation of the stimulus but an endogenously generated *response triggered by the sensory perturbation* [...] (Thompson 2007, p. 368 italics added)

If autonomy is to be reconciled with sensorimotor enactivism, it is through the fundamental circularity that grounds sensorimotor autonomy: sensorimotor correlations depend on the neurodynamic patterns (and body-environment structures) that bring them about while the neurodynamic patterns depend on their creation and maintenance on the sensorimotor correlations they bring about. By acknowledging this circularity we have also overcome the problem of non-intersecting domains: sensorimotor contingencies are not relational, observer-dependent and operationally irrelevant or ineffective correlations, but dynamically required constraints, direct (non-mediated), inalienable constituents of the operational closure of the “nervous + sensorimotor” systems.

## 4.2 Sensorimotor Autonomy Revisited

The sensorimotor constitution of neurodynamic patterns is not the only relevant aspect of autonomy that matters to enactivism. In fact, we are still far from a theory of sensorimotor autonomy. What this theory needs to provide is the satisfaction of the three necessary and sufficient conditions for the emergence of autonomous agency at the scale of sensorimotor coordination dynamics.<sup>14</sup> We can properly talk of agency when: “(a) there is a system as a *distinguishable entity* that is different from its environment [individuality condition], (b) this system is *doing* something by itself in that environment [interactional asymmetry condition], and (c) it does so according to a certain *goal* or *norm* [normativity condition]” (Barandiaran et al. 2009, p. 369). What enactivism hasn’t yet been able to make explicit—despite some theoretical approximations (Barandiaran 2008; Di Paolo 2003, 2005)—is how exactly sensorimotor identity or individuality, interactional asymmetry and the origin of cognitive or behavioural norms can be naturalized in terms of a sensorimotorly constituted neurodynamic organization.

We need to move beyond the case of a single neurodynamic pattern to an *organization* of such patterns *in* interaction with the environment. The notion of habit might prove helpful, at this point, as a building block for a theory

<sup>14</sup> Agency does not here imply necessarily the category of personal against that of the subpersonal. It does however imply the emergence of an interactive, i.e. sensorimotorly constituted, unit or identity understood as a whole (in a manner that is not reducible to the workings of specific neural, bodily or environmental mechanisms taken in isolation).

of cognitive organization. The more familiar and synthetic concept of habit provides, in a nutshell, an intuitive grasp that is closer to mainstream philosophical theorizing than what complex models of non-linear dynamics can provide. Cognitivism did its best to definitely debunk a notion of habit that was itself a *redux* version of what behaviourism made acceptable through the filter of operationalist epistemology applied to associationism (ignoring the bodily and neural generative mechanisms that were far from scientific operational modelling at the time). But an alternative conception of habit can be recovered from a rich and inspiring history that can be traced back from Aristotle’s ethics to Piaget through an organicist school of thinking of which enactivism is itself an heir (Barandiaran and Di Paolo 2014). Beyond the mere stimulus-triggered response probability, this richer conception of habit integrates brain, body and sensorimotor dynamics into a self-sustaining behavioural “life-form” (Egbert and Barandiaran 2014).

As a first approximation, we can re-define habit as “a *self-sustaining pattern of sensorimotor coordination* that is formed when the stability of a particular mode of sensorimotor engagement is dynamically coupled with the stability of the mechanisms generating it” (Barandiaran 2008, p. 281).<sup>15</sup> What habits, thus understood, add to the sensorimotor correlations and neurodynamic patterns we have previously seen is the notion of plasticity: the fact that repeated enactments of a given sensorimotor correlation pattern bear plastic reinforcing consequences on the mechanisms (e.g. synaptic branching) that structurally support it.

A single habit provides a first analogy with life and a first approximation to a sensorimotor conception of identity and normativity (Egbert and Barandiaran 2014). Through repetition (and the myriad of reinforcing plastic mechanisms that brain and world can provide) a habit can take on a life of its own: it is both the cause and the consequence of its own enactment. This form of recursion makes it possible to understand a mild sense of identity for the habit, a locus of survival and self-generating persistence—as Ravaïsson so cleverly anticipated almost three centuries ago

<sup>15</sup> In Egbert and Barandiaran (2014) we developed a robot controller that instantiates this concept of habit. We embedded the robot with a *deformable sensorimotor medium* that mediated between sensory and motor surfaces, a continuous transfer function that can be depicted as multidimensional vector space (the dimensions being the sensory and motor variables). We designed the system so that trajectories along the sensorimotor medium were self-reinforcing: i.e. the weight of the vectors increased each time a trajectory was made in the sensorimotor space. Through repetition, and provided that the sensorimotor environment makes possible the enactment of a particular sensorimotor trajectory, paths are carved out in the deformable medium. The result is a complex topology of traces and recurrent shapes that fade away if the agent-environment system fails to re-enact, and get reinforced through repetition.

(Carlisle 2014; Ravaissón 2008). The precarious dependence of habitual behaviour on the sensorimotor environment and brain-body structures defines a set of viability conditions that impose certain normative constraints, those that ensure the “survival” of the habit: the environmental (sensorimotor regularity) conditions within which it can recur, the required rhythm or repetition, the range of pattern variations within which the habit reinforces without splinting or vanishing, etc. i.e. the conditions under which the habit is kept “alive”.

This is however a very first approximation, we need to go beyond a single self-reinforcing habit to get closer to a notion of sensorimotor autonomous agency. William James conceived of animals as “bundles of habits”. This notion of a bundle, or an ecology of habits (with habits competing for sensorimotor enactment, or cooperating, nesting, sequencing, etc.), both makes possible and depends on a sensorimotor environment or *habitat*. Out of an undifferentiated sensorimotor surrounding (defined by the potential set of sensorimotor contingencies that the body of the agent could explore, given the spatio-temporal scale, accuracy and articulation of its sensors, motors, etc.) the bundle co-determines a subset of sensorimotor contingencies that it *inhabits*. The agent’s effective environment (its habitat) is cut out of its potential environment, by habitually avoiding some correlations (e.g. approaching dark corners), selecting others, actively seeking some, sequentially switching from one to another, etc. (see Buhrmann et al. 2013). This bundle of habits is, partly, tied or meshed within the brain, where most of the reliable plasticity of habits and the capacity for selective enactment lies.

Within a relatively complex brain, the self-maintenance of habits, or neurodynamic structures, need not be reduced to simple recurrent self-reinforcement. Stability dependencies, transitions, co-activations, etc. might rely, not on isolated habitual sensorimotor structures, but on more relationally complex, interdependent architectures. Anatomical constraints and activity-dependent plasticity in the brain, in combination with a rich environment and an active history of developmental scaffolding, affords for quite a complex organization of habits or sensorimotor structures (Johnson 2001).

If the level of plastic interconnectedness of these bundles is complex enough, at some point of its development, it is reasonable to assume that sensorimotor regulation might not be driven exclusively by the dominance of a single habit (or, to put it in neurological terms, by the influence of a unique anatomically individualized pathway or an isolated synchrony of oscillatory neuronal assembly). Sensorimotor regulation will soon give rise to large-scale equilibrating tensions within the bundle. We can add that a proper sense of autonomous sensorimotor agency, or *Mental Life*, comes into existence when the adaptive

conservation of this bundle becomes the main principle of sensorimotor regulation; when sensorimotor compensations, accommodations, re-arrangements, etc. take place to maintain the capacity of the agent to keep behaving coherently—for a more detailed account of these issues see Barandiaran (2008, Chapter 8).<sup>16</sup> A new normative domain emerges: e.g. the *failure* of the environment to collaborate in the balancing of conflicting habits or the difficulty to *appropriately* enact the *right* sensorimotor coordinations on which a tangle of habits depends for its systemic equilibrium. This can happen due, for instance, to the presence of visual inversion goggles (Kohler 1964), the *maladaptive* assimilation of tools that disrupts a habitual task, the *skilful* recovery of performance after a breakdown, the *successful* accommodation to a prosthesis, etc.

The detailed neuroscientific, ecological and behavioural support for this picture is outside of the scope of this paper, but I would like to highlight two points that anticipate some objections and clarify potential misunderstandings. First, despite the cognitivist cornering of an already impoverished version of behaviorists’ habit, what is still commonly accepted as pure habit in neuroscience falls acknowledgedly far from simple stimulus-response circuits in the basal ganglia. Neuroscientists acknowledge the need to develop dynamic models of the simultaneous and distributed activity patterns that support habits in cortico-basal ganglia loops (Graybiel 2008, pp. 377–378). Moreover, and this brings us to the second point, neurodynamic structures of the kind just hypothesized to constitute sensorimotor autonomy need not be reduced to the neurobiology of what is currently understood as “habitual behaviour”. Large-scale neuroscience provides a stock of theoretical building blocks that satisfy the demands for a theory of behavioural neurodynamic autonomy: from neuronal *cell assemblies* (Hebb 1949; Varela 1995) or *dynamic cell assemblies* (Tsukada et al. 1996) to the *dynamic core* hypothesis (de Pasquale et al. 2012; Edelman and Tononi 2000), from *global attractors* (Freeman 2001) to *dissipative dynamic structure* (Llinás 2001), from *chaotic attractors* (Tsuda 2001) to the *effective connectivity of distributed neuronal ensembles* (Fries 2005), etc. The dynamic participation of motor regions (previously thought to be physiologically decoupled or epistemically separable from “cognitive

<sup>16</sup> The autonomy of this sensorimotor domain, does not imply independence from metabolic or physiological needs of the organism. The enbrainment of the body makes for an intimate bidirectional influence between sensorimotor or cognitive and biological norms and processes. Biological needs might manifest themselves, to the autonomous web of neurodynamic structures, as endogenously originated perturbations to the stability of this habit-ecology (e.g. the urge to find food or water).

processing") in the formation of such patterns is nowadays being increasingly acknowledged (Engel et al. 2013).<sup>17</sup>

The degree to which these large-scale distributed neurodynamic patterns interrelate so as to determine collective or global-organizational viability conditions (i.e. norms) in continuous embodied interaction with the environment is hard to determine in detail. Some progress in systems-neuroscience and consciousness could shed light on this matter, but the philosophical dispute can already be settled in or at least can be sufficiently framed in terms of non-representational dynamic sensorimotor *coherence* (as anticipated by Varela 1992, 1997). Framing this notion of coherentism within sensorimotor contingency theory (or within some forms of enactivism) might come as a novel contribution but it is not, in itself, a new idea:

Through learning, a complex schema network arises that can mediate first the child's, and then the adult's, reality. Through being rooted in such a network, schemas are interdependent, so that each finds meaning only in relation to others. (...) Each schema enriches and is defined by the others (...). Though processes of schema change may affect only a few schemas at any time, such changes may "cohere" to yield dramatic changes in the overall pattern of mental organization. (Arbib et al. 1998, p. 44)

At multiple levels of analysis at multiple time-scales, many components open to influence from the external world interact and in so doing yield coherent higher-order behavioural forms that then feedback on the system, and change that system. (Smith and Thelen 2003)

The type of neuro-sensorimotor coherentism that I am proposing here avoids the solipsism often attributed to Maturana and Varela's work (Mingers 1994). What I defend is a notion of autonomy that centres a perspective and co-defines a world that is *constitutively* sensorimotor. Neurodynamic coordination patterns generate, through body and environment, sensorimotor coordinations, that in turn support and make possible functionally specific neurodynamic patterns. At the same time, at a slower timescale, these patterns shape, reinforce and re-structure the plastic neural organization and the habitat or niche construction that gives rise to them. The operational closure should not be understood in terms of "plain" neuronal activity (as proposed by the early Maturana and Varela), but, rather (as suggested by Varela 1992, 1997), at the higher level of patterns of coordinated neuronal activity and sensorimotor coupling taken together. We can illustrate

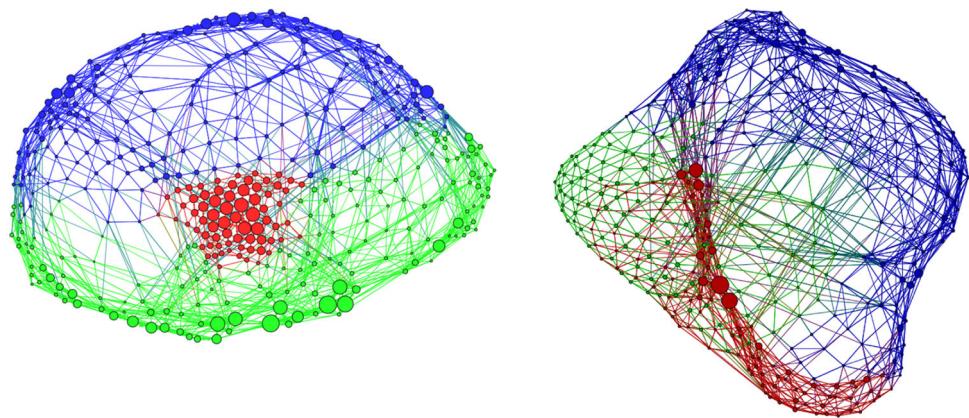
and naturalize some aspects of this sensorimotor neurodynamic autonomy with a simulation model.

In (Aguilera et al. 2015), and inspired by previous work by Iizuka and Di Paolo (2007), we explored a minimally complex neurodynamic organization that involves a network of emerging patterns of oscillatory coordination and plasticity in situated activity. The robot spontaneously alternates between two behavioural preferences: approaching either one of two light sources in a two dimensional arena. The robot controller is composed of three Kuramoto oscillators that coordinate their behaviour entering into different relative-phase patterns during each phototactic episode. Activity dependent plasticity transforms the network during such episodes thus changing the resulting patterns. Figure 3 (left) shows the network of the oscillatory neurodynamic patterns that emerge when the agent is coupled to the environment: each node of the network corresponds to a pattern and the edges of the graph correspond to the transitions between these patterns. The network is drawn for a long history of behavioural choices. Nodes were coloured according to a graph-theoretical modularity algorithm that identifies partitions in the network. Blue and green nodes correspond to phototactic committed preference to blue and green light sources respectively, and red corresponds to episodes in which the agent's behavioural preference is "uncommitted" and open to environmental influence to approach either one of the two light sources.<sup>18</sup>

It is at the level of this abstract network that operational closure is to be pictured. Both the circles (patterns) and the transition network are constitutively sensorimotor (in the sense outlined in the previous section). We carried out the same experiment with a twin robot fed with the very same sensory input-stream. The closed network that results from the freely behaving agent compared to the partially-coupled agent (right illustration in Fig. 3) is not only composed of different patterns, but also displays a different topology. Structural (sensorimotor) coupling is not something that happens on top of an already closed network. The operational closure should not be specified at the level of a neuronal network pictured as an ensemble of interconnected neurons (as Maturana and Varela often seem to imply), but, contrary to Bittorio, as a *network of transitions between neurodynamic patterns that are (both the patterns and the transition) constitutively sensorimotor*. The network depicted in Fig. 3 is generated from a 3 oscillator network, yet it displays dozens of nodes (oscillatory

<sup>17</sup> An empirical development of the sensorimotor constitution thesis would demand that  $M \rightarrow S$  correlations become essential to the modelling of such patterns, assemblies or structures.

<sup>18</sup> Although the images of Fig. 3 are not included in the original paper, specific details of this model and its interpretation in terms of behavioural sensorimotor autonomy can be found at (Aguilera 2015), together with fractal analysis of the sensorimotor coupling and the systematic study of the role of plasticity to generate self-organized behavioural criticality.



**Fig. 3** Oscillatory neurodynamic patterns (circles) connected through transition diagrams, forming a network, for an homeostatically plastic 3 oscillator network, controlling a robot that alternates between phototaxis to two different light sources (green and blue), red nodes represent neurodynamic patterns that corresponds to non committed agent (whose behavioural preference is not internally fixed, but open to environmental variations), blue nodes correspond to the agent committed to approaching the blue light source and green

patterns) and edges (transition links between patterns). The robustness of the network and the role of plasticity to reshape and regulate the topology through interactions can provide formal (unambiguous and operational) criteria to depict coherency and the emergence of norms as the equilibrating conditions of an ecology of patterns (or habits).

Following the analogy with biological life, mental death occurs when continuous disruption of the sensorimotor coupling irreversibly destroys the capacity of the system to behave coherently, i.e. the organization of behaviour is lost. In the model we just introduced, if the network topology irreparably loses its structure, if coordination patterns are lost (nodes disappear), potential transitions are weakened and perish (edges or links disappear) autonomous sensorimotor agency will be gone, the system will lose its capacity to behave. Thus a specific (sensorimotor) normative dimension can be operationalized or measured for this kind of network of habits: the viable limits (of disruption, decoupling, etc.) out of which the organization is irreversibly lost. A norm emerges, taking the form of a Kantian imperative or regulatory principle: *behave so as to sustain your capacity to behave*.

## 5 Discussion: Enactivism Without Subject, Without Norms, Without Meaning?

We have seen how a notion of autonomy that focuses exclusively on autopoiesis (or metabolic closure) falls short of providing a theorizing ground for sensorimotor enactivism. Moreover, the original formulation of autonomy in terms of the operational closure of the nervous system has

nodes to a behavioural preference towards *green light sources*. The network on the left corresponds to a freely behaving agent, whereas the network on the right corresponds to an passively-coupled agent receiving the same exact input as the first but without this input being the result of its own movement. [Unpublished results from (Aguilera et al. 2015), available under a Creative Commons By-SA license at <http://maguilera.net>]. (Color figure online)

been shown to be deeply problematic when combined with the “sensorimotor constitution of experience” thesis. As a result, autonomy became a concept that was hard to assume and relatively simple to ignore. It is therefore understandable that the presence of autonomy as a theoretically useful notion has vanished within some recent enactivist trends. But, as we make progress on constructive theorizing, enactivists cannot afford to ignore (or to reject) what theories of autonomy have to offer.

We have shown that a sense of sensorimotorly constituted autonomous agency is perfectly compatible (and complementary) with those varieties of enactivism that have ignored or downplayed what autonomy has to offer as a foundational concept for enactivism. Conceptual simulation models, as guides to discovery, epistemic tools or philosophical thought experiments (Barandiaran and Chemero 2009; Barandiaran and Moreno 2006a; Di Paolo 2000) provide the means to illustrate, distil and conceptually sharpen the complex dynamic relationships that this notion of autonomy demands. There is certainly still much progress to be made to deliver a proper dynamical ontology that re-constructs cognitive categories (perception, intention, action, learning, imagining, etc.) in terms of autonomist enactivism (although some advances are already available Buhrmann et al. 2013; Di Paolo et al. 2014). Meanwhile we can benefit from the theoretical progress achieved so far to discuss some of the implications that it bears within contemporary debates on sensorimotor or embodied cognitive science.

To start with, if we ignore the gap that the notion of autonomy comes to fill within enactivism we are left with a research programme where missile seeking devices and

bacteria alike—whatever might come to be described as being “attuned to the ways in which one’s movements will affect the character of input” (O’Regan and Noë 2001, p. 84)—can be considered cognitive. To strengthen the requirement to “knowledge of sensorimotor contingencies” won’t solve the problem, since it leaves open precisely the central question that a research programme in cognitive science needs to define: who is the subject of cognition (who is the bearer or the concerned with knowledge) and which are the principles that set up the normative framework for the deployment of that knowledge. No matter how skilful or non-intellectualist is the conception of knowledge, it demands to be subsumed under norms: to be right or wrong, adaptive or maladaptive, clumsy or skilful, joyful or frustrating, satisfying or uncongenial.<sup>19</sup>

More recently, autonomy, in the worldly and normative dimension I have defended here, has been put into question within enactivism (Hutto and Myin 2012) and conservative Maturanian approaches (Villalobos 2013). For both, autonomist enactivism risks sliding into representationalism by defending norms, values or meaning.

A physicalist-mechanist perspective, of the type advocated for in Maturanian conservatism (Villalobos 2013), one that hopes to remove any reference to norms, value, teleology, function or meaning, is unsustainable, even within the “mechanistic” framework that Maturana defends. Villalobos quotes Maturana in a critical passage for sensorimotor enactivism:

In the organization of the living systems the role of the effector surfaces [with or without a nervous system] is only to maintain constant the set states of the receptor surfaces, not to act upon an environment, no matter how adequate such a description may seem to be for the analysis of adaptation, or other processes (Maturana 1970/Maturana and Varela 1980, p. 51)

The first point to be noted is that Maturana uses the teleological term “role”. But physical laws or deterministic differential equations simply have no roles. The concept of role implies an assignment of a specific function (within a normative context) to a structure or process that could (potentially) operate otherwise. This is not simply a contingent terminological accident. Teleological vocabulary

steps, again, into Maturana’s explanation when using the notion “to maintain”, for maintaining has to do with correcting deviations. The crucial point here is that one can *fail* to maintain. A raw physicalist-mechanistic framework, however, is unable to deliver, on its own, any sense of norms that can justify such failure. Very simply put, if the *role* of X is to *maintain* Y, then we are not giving a deterministic description at all. The alternative is simply to say that X does Y, and this is false simply if X doesn’t do Y. But, obviously, this is often (most often) the case: effector surfaces do not (always) “maintain constant the set of states of the receptor surfaces”,<sup>20</sup> living systems do not always succeed. But they try, meaning that there exist regulatory principles and self-organizing tendencies that result from large scale organizational configurations. And these are essential to characterize cognition and cognitive processes. Mechanistic explanations, in biological and cognitive sciences, are inescapably framed within the organizational context of the whole organism and within its autonomy (Bechtel 2007a, b; Levy and Bechtel 2013; Moreno et al. 2011).

A less mechanistic critique of autonomy comes from Radical Enactivism’s main thesis (or counter-thesis as the authors want to put it) stating that “there can be intentionally directed cognition and, even, perceptual experience without content” (Hutto and Myin 2012, p. 10), and they further specify that “there is content wherever there are specified conditions of satisfaction” (Hutto and Myin 2012, p. 10). Following a Hegelian impulse, I suggest here a negation of the negation, a counter-counter-thesis: there cannot be intentionally directed process without “specified conditions of satisfaction”. Hutto and Myin did an excellent job debunking representational content as a useful explanatory resource for cognitive science but they don’t provide an alternative theory of intentionality. The Developmental-Explanatory Thesis (Hutto and Myin 2012) won’t do the explanatory job: “The secret to explaining what structures an organism’s current mental activity lies entirely in its history of previous engagements and not in some set of internally stored mental rules and representations” (p. 9). But why should conditions of satisfaction be stored as mental rules or representations? Autonomy fills the gap that Hutto and Myin assume to be insurmountable between a historical perspective and representationalism. It is no secret but a physical fact, of which Varela et al. were perfectly aware, that history needs a subject (a character that is both cause and effect of its own history) to leave its traces, to be explanatory of anything; that no plain

<sup>19</sup> A similar, and much more detailed, criticism to Noë, O’Regan and Hurley’s approaches is made by Evan Thompson (2006, 2007). My contribution expands and complements his criticism and alternative proposal. Thompson develops the enactivism position primarily in terms of the first-person phenomenology of the lived body. But his account of sensorimotor autonomy falls short of solving the problems I outlined throughout Sect. 3 and he provides no explicit formulation, or models, that redefine autonomy in terms of sensorimotorly constituted neurodynamics (as that provided in Sect. 4 of this paper).

<sup>20</sup> The argument equally works if the role of the effector surfaces is to generate a particular sensory structure (as enactivism would imply) and not to maintain it stable, an idea magnificently developed by Powers (1973).

historical record of interactive facts can substitute the actual mechanisms that are producing behaviour within the organism in continuous embodied interaction with its environment. To which extent the historical traces need to be located inside the organisms or within its environment is open to specific empirical investigation, but history doesn't explain without a subject, and without the world it carves out of its sensorimotor environment, without specifying the kinds of mechanisms and organizational processes involved in its ongoing cognitive coupling. What autonomist enactivism emphasizes is precisely that there is no explanatory completeness if we don't combine the dynamic study of the sensorimotor coupling with the embodied neurodynamic organization that sustains and produces it (and vice versa). And the specification of this subject and its behavioural organization does not have to be representational, there are sufficient theoretical resources in place to escape the exclusive choice between either history or representationalism.

Hutto and Myin's critique of autopoietic enactivism becomes explicit when approaching the notions of meaning and value (that they equate with that of content). Although some authors seem to imply it, by an often obscure and imprecise use of the terms "meaning" and "value" (see Hutto and Myin 2012, pp. 33–36), the concept of autonomy does not necessarily imply that meaning is a commodity, a content to be carried out by vehicles (to be extended into body and environment). *Meaning is use value for the organism*, and not in the form of a signal, a trigger or a environmental perturbation, but the "use value" of a form of sensorimotor engagement for the satisfaction of equilibrating conditions of a precarious organization of habits. Habits that, as we have just seen, rely on the enactment of sensorimotor contingency structures (the habitat) for their collective self-maintenance. If, and only if, the neurodynamic structures that mediate such an environment are organizationally precarious, co-dependent on the right sensorimotor contingencies enacted through action, does a sense of meaning become available: a normative tendency to assimilate environmental variations or accommodate the organization of mediating neurodynamic structures to changing or novel conditions.<sup>21</sup>

What autonomy provides is that these conditions of satisfaction are specified by the very system, and don't take the

form of vehicles of any kind aimed at encapsulating informational or representational content. The world is not different from the experience, from the enactment of the sensorimotor coordinations that make certain neurodynamic patterns possible. This is why there is no content-of or content-about the world. And yet the world is meaningful. In its minimal form it appears with gradients of attraction or repulsion; e.g. moving the agent away from those modes of engagement that threaten the viability of the system. More complex modes of agency might include multidimensional gradations of the adaptive value of certain sensorimotor enactments. Rejection of representational content does not imply rejection of value or norm, of sense, of meaning.

A naturalized account of norms (or normative functionality) in terms of the viability constraints of an autonomous organization has been developed at length in biology, mostly in opposition to evolutionary or adaptationist accounts of normative function (Barandiaran and Egbert 2013; Barandiaran and Moreno 2008; Christensen and Bickhard 2002; Mossio et al. 2009). We have proposed that this account of normativity can be relatively easily transferred, by analogy, to mental life, through the notion of a self-sustaining network of habits and its coherentist dynamical demands. The emergence of this level of autonomous agency has important consequences that make possible to declare it as minimally cognitive: it grounds a worldly normativity, based on sensorimotor dynamics. This level of normativity is not reducible to biological normativity, neither does it demand socio-linguistic skills or embedding—as recently argued against enactivism (Heras-Escribano et al. 2014). And yet, it remains open to both. On the side of biological embodiment sensorimotor dynamics are embedded in an organismic whole whose physiological functions are also innervated and interplay with sensorimotor dynamics. On the side of social embodiment mental life is, for many species, constitutively social, open to the coupling with other agents whose collective social dynamics shape the neurodynamic organization of behaviour (De Jaegher et al. 2010). The way in which social and biological dynamics constrain, amplify, bootstrap, and channel sensorimotor dynamics is critical to the formation of the human mind, but there is no need to resort to social interaction to ground a naturalized organizational conception of sensorimotor normativity. Moreover, this notion of normativity escapes the non-teleological and reductive conception of the operational closure of the nervous system (Villalobos 2013), and it grounds the anti-representationalist claims in a productive alternative. These are crucial features within the contemporary enactivist context. If we ignore the contribution that theories of autonomy have to make for enactivism we risk throwing the baby out with the bathwater, or throwing meaning out with representationalism.

<sup>21</sup> I tend to avoid the use of "meaning" and "sense making" myself in favour of normativity. Although, in most cases, once a normative dimension of behaviour is in place, "meaning" and "sense making" can be re-phrased in terms of the valence of a normatively framed interaction, I prefer to be cautious and save these terms for higher, more complex, forms of agency that might demand linguistically scaffolded forms of cognitive autonomy or specific forms of emotional shaping of sensorimotor dynamics (Barandiaran 2008, Chapter 8.5.3).

A final objection to autonomist enactivism, made by Hutto and Myin, has to do with the phenomenological approach that is often combined with autonomy, embodiment and sensorimotor coupling (Thompson 2007; Varela et al. 1991; Weber and Varela 2002). It is important to note that there is no necessary dependency of autonomist enactivism upon existentialism or phenomenology. Theory construction (the systematic accumulation of definitions, experimental procedures, models, etc.) can be carried out in perfectly naturalized operational terms (i.e. in a manner in which all terms and hypothesis can be articulated by means of explicit manipulation or operations to be carried out in experimental set-ups or in formal or computational models). Phenomenology can certainly provide a guide to discovery and some kind of experiential test bed or confirmation. Phenomenology can be used to ensure that a set of operational distinctions have been made in congruence with the structure of our lived experience. But once the phenomenological delimitation has been fixed and the formal definition adjusted to it, the procedure to identify or measure such a phenomenon, to model it, or to artificially create one, is independent. More technically speaking, phenomenology can provide a *context of discovery*, while dynamical systems theory or other formalisms and conceptual resources, together with empirical data and modelling techniques, provide for a *context of justification*. In other words, phenomenology can move the leg to provide the heuristics of where the joints of cognitive reality are; while science (formalization, modelling, experimentation) provides the knife to chop it right at the joints. What matters to the scientific quality of the result is the chopping procedure (the knife cut). Whether the heuristic is phenomenological (or folk-psychological) should not bother the epistemic hygienist, but it matters to the butcher that actually does the job and so it is reasonable to assume a role for phenomenology (or philosophy more generally) in embodied and autonomist approaches to enactivism.

All these questions are not mere scholastic disputes within a relatively new theoretical framework that struggles to set up a consensus on its research agenda. On the contrary, the issue of autonomy is central to cognitive science and philosophy of mind itself. The specificity and autonomy of cognition is at stake, the nature of normativity, the basic epistemological and metaphysical principles that articulate scientific progress. Autonomy emphasizes the self-organized, holistic, dynamic interdependence within self-sustaining organizations, it challenges representationalist realism as a way to approach agent-environment relationships by highlighting the dialectic co-dependence between the identity of a system and the habitat it selects, shapes and brings about through its specific mode of coupling. Autonomy anchors normativity in the large-scale plastic correlations and homeostatic inter-

dependencies of sensorimotor coherency, instead of relying on linguistic, evolutionary/adaptationist or representationalist principles. Cognitive science, enactivist or otherwise, cannot afford to disregard the question of how the subject of experience emerges from sensorimotor interactions and how it operates according to the norms that such interactions bring forth. Autonomist sensorimotor enactivism provides a research agenda to address these pressing questions.

**Acknowledgments** This work was partially funded by the eSMCs (Extending Sensorimotor Contingencies to Cognition) project (ref. FP7-ICT-2009-6, IST-270212) founded by the EU, research project “Autonomía y Niveles de Organización” financed by the Spanish Government (ref. FFI2011-25665) and IAS-Research group funding IT590-13 from the Basque Government, and the project “Identidad en interacción” (ref. FFI2014-52173-P) founded by MINECO, Spanish Government.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

Adams S (2007) Castoriadis and autopoesis. *Thesis Eleven* 88(1):76–91. doi:[10.1177/0725513607072458](https://doi.org/10.1177/0725513607072458)

Adams F, Aizawa K (2009) Why the mind is still in the head. In: Robbins P, Aydede M (eds), *The Cambridge Handbook of Situated Cognition*. Cambridge University Press, pp. 78–95

Aguilera M (2015) Interaction dynamics and autonomy in cognitive systems, from sensorimotor coordination to collective action (PhD Thesis). Universidad de Zaragoza, Zaragoza, Spain. <http://maguilera.net/phdthesis>

Aguilera M, Bedia MG, Santos BA, Barandiaran XE (2013) The situated HKB model: how sensorimotor spatial coupling can alter oscillatory brain dynamics. *Front Computat Neurosci*. doi:[10.3389/fncom.2013.00117](https://doi.org/10.3389/fncom.2013.00117)

Aguilera M, Barandiaran XE, Bedia MG, Seron F (2015) Self-organized criticality, plasticity and sensorimotor coupling. Explorations with a neurorobotic model in a behavioural preference task. *PLoS One* 10(2):e0117465. doi:[10.1371/journal.pone.0117465](https://doi.org/10.1371/journal.pone.0117465)

Aizawa K (2007) Understanding the embodiment of perception. *J Philos* 104(1):5–25

Arbib MA, Érdi P, Szentágothai J (1998) Neural organization: structure, function, and dynamics. MIT Press, Cambridge

Arnellos A, Moreno A, Ruiz-Mirazo K (2014) Organizational requirements for multicellular autonomy: insights from a comparative case study. *Biol Philos* 29(6):851–884. doi:[10.1007/s10539-013-9387-x](https://doi.org/10.1007/s10539-013-9387-x)

Barandiaran XE (2004) Behavioral adaptive autonomy. A milestone on the Alife route to AI? In: Pollack J, Bedau MA, Husbands P, Ikegami T, Watson RA (eds) *Artificial life IX: proceedings of the ninth international conference on the simulation and synthesis of artificial life*. MIT Press, Cambridge, pp 514–521

Barandiaran XE (2007) Mental Life: conceptual models and synthetic methodologies for a post-cognitivist psychology. In: Wallace B, Ross A, Davies J, Anderson T (eds) *The world, the mind and the*

body: psychology after cognitivism. Imprint Academic, Exeter, pp 49–90

Barandiaran XE (2008) Mental Life: a naturalized approach to the autonomy of cognitive agents. (PhD Thesis). University of the Basque Country (UPV-EHU), Donostia - San Sebastián, Gipuzkoa, Spain. <http://barandiaran.net/phdthesis/>

Barandiaran XE, Chemero A (2009) Animats in the modeling ecosystem. *Adapt Behav* 17(4):287–292. doi:[10.1177/1059712309340847](https://doi.org/10.1177/1059712309340847)

Barandiaran XE, Di Paolo EA (2014) A genealogical map of the concept of habit. *Front Hum Neurosci* 8:522. doi:[10.3389/fnhum.2014.00522](https://doi.org/10.3389/fnhum.2014.00522)

Barandiaran XE, Egbert MD (2013) Norm-establishing and norm-following in autonomous agency. *Artif Life* 19(2):1–24. doi:[10.1162/ARTL\\_a.00094](https://doi.org/10.1162/ARTL_a.00094)

Barandiaran XE, Moreno A (2006a) ALife models as epistemic artefacts. In: Rocha L, Yaeger L, Bedau M, Floreano D, Goldstone R, Vespignani A (eds) *Artificial Life X*: proceedings of the tenth international conference on the simulation and synthesis of living systems. The MIT Press (Bradford Books), Cambridge, pp 513–519

Barandiaran XE, Moreno A (2006b) On what makes certain dynamical systems cognitive: a minimally cognitive organization program. *Adapt Behav* 14(2):171–185. doi:[10.1177/105971230601400208](https://doi.org/10.1177/105971230601400208)

Barandiaran XE, Moreno A (2008) Adaptivity: from metabolism to behavior. *Adapt Behav* 16(5):325–344. doi:[10.1177/1059712308093868](https://doi.org/10.1177/1059712308093868)

Barandiaran XE, Ruiz-Mirazo K (2008) Modelling autonomy: simulating the essence of life and cognition. *Biosystems* 91(2):295–304. doi:[10.1016/j.biosystems.2007.07.001](https://doi.org/10.1016/j.biosystems.2007.07.001)

Barandiaran XE, Di Paolo E, Rohde M (2009) Defining agency: individuality, normativity, asymmetry, and spatio-temporality in action. *Adapt Behav* 17(5):367–386. doi:[10.1177/1059712309343819](https://doi.org/10.1177/1059712309343819)

Bechtel W (2007a) Biological mechanisms: organized to maintain autonomy. In: Boogerd FC, Bruggeman FJ, Hofmeyr JHS, Westerhoff HV (eds), *Systems Biology: Philosophical Foundations*. Elsevier, Amsterdam, The Netherlands, pp. 269–302

Bechtel W (2007b) Mental mechanisms: philosophical perspectives on cognitive neuroscience, 1st edn. Psychology Press, Hillsdale

Di Paolo EA, Noble J, Bullock S (2000) Simulation models as opaque thought experiments. In: Bedau M, McCaskill J, Packard N, Rasmussen S (eds) *Artificial Life VII*: proceedings of the seventh international conference on artificial life. MIT Press, pp 497–506

Bickhard MH (2000) Autonomy, function, and representation. *Commun Cogn Artif Intell* 17(3–4):111–131

Bickhard MH (2009) The biological foundations of cognitive science. *New Ideas Psychol* 27(1):75–84

Block N (2005) Review of Alva Noë, action in perception. *J Philos CII*(5):259–272

Bompas A, O'Regan JK (2006) Evidence for a role of action in colour perception. *Perception* 35(1):65–78. doi:[10.1068/p5356](https://doi.org/10.1068/p5356)

Bourgine P, Stewart J (2004) Autopoiesis and cognition. *Artif Life* 10(3):327–345

Buhrmann T, Di Paolo EA, Barandiaran XE (2013) A dynamical systems account of sensorimotor contingencies. *Front Cogn 4*:285. doi:[10.3389/fpsyg.2013.00285](https://doi.org/10.3389/fpsyg.2013.00285)

Burge T (2009) Primitive agency and natural norms. *Philos Phenomenol Res* 79(2):251–278. doi:[10.1111/j.1933-1592.2009.00278.x](https://doi.org/10.1111/j.1933-1592.2009.00278.x)

Carlisle C (2014) On habit. Routledge, London

Castoriadis C (1997) *World in fragments: writings on politics, society, psychoanalysis, and the imagination*. Stanford University Press, Redwood City

Chemero A (2009) Radical embodied cognitive science. The MIT Press, Cambridge

Christensen WD, Bickhard MH (2002) The process dynamics of normative function. *The Monist* 85(1):3–29

Christensen WD, Hooker CA (2000) Autonomy and the emergence of intelligence: organised interactive construction. *Commun Cogn 17*(3–4):133–157

De Jaegher H, Di Paolo E (2007) Participatory sense-making. *Phenomenol Cogn Sci* 6(4):485–507

De Jaegher H, Di Paolo EA, Gallagher S (2010) Can social interaction constitute social cognition? *Trends Cogn Sci* 14(10):441–447. doi:[10.1016/j.tics.2010.06.009](https://doi.org/10.1016/j.tics.2010.06.009)

de Pasquale F, Della Penna S, Snyder AZ, Marzetti L, Pizzella V, Romani GL, Corbetta M (2012) A cortical core for dynamic integration of functional networks in the resting human brain. *Neuron* 74(4):753–764. doi:[10.1016/j.neuron.2012.03.031](https://doi.org/10.1016/j.neuron.2012.03.031)

Di Paolo EA (2000) Homeostatic Adaptation to Inversion of the Visual Field and Other Sensorimotor Disruptions. In: Meyer JA, Berthoz A, Floreano D, Roitblat H, Wilson SW (eds) *From animals to animats 6*. MIT Press, Cambridge, MA, pp. 440–449

Di Paolo EA (2003) Organismically inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop. In: Murase K, Asakura T (eds) *Dynamical systems approaches to embodiment and sociality*. Advanced Knowledge International, Adelaide, pp 19–42

Di Paolo EA (2004) Unbinding biological autonomy: Francisco Varela's contributions to artificial life. *Artif Life* 10(3):231–233

Di Paolo EA (2005) Autopoiesis, adaptivity, teleology, agency. *Phenomenol Cogn Sci* 4(4):429–452. doi:[10.1007/s11097-005-9002-y](https://doi.org/10.1007/s11097-005-9002-y)

Di Paolo EA, Thompson E (2014) The enactive approach. In: Shapiro L (ed) *The Routledge handbook of embodied cognition*. Routledge, London, pp 68–78

Di Paolo EA, Barandiaran XE, Beaton M, Buhrmann T (2014) Learning to perceive in the sensorimotor approach: Piaget's theory of equilibration interpreted dynamically. *Front Hum Neurosci* 8:551. doi:[10.3389/fnhum.2014.00551](https://doi.org/10.3389/fnhum.2014.00551)

Edelman GM, Tononi G (2000) *A universe of consciousness: how matter becomes imagination*. Basic Books, New York

Egbert MD, Barandiaran XE (2014) Modeling habits as self-sustaining patterns of sensorimotor behavior. *Front Hum Neurosci* 8:590. doi:[10.3389/fnhum.2014.00590](https://doi.org/10.3389/fnhum.2014.00590)

Engel AK (2010) Directive minds: how dynamics shape cognition. In: Stewart JR, Gepenne O, Di Paolo EA (eds) *Enaction: toward a new paradigm for cognitive science*. MIT Press, Cambridge, pp 219–265

Engel AK, Maye A, Kurthen M, König P (2013) Where's the action? The pragmatic turn in cognitive science. *Trends Cogn Sci* 17(5):202–209. doi:[10.1016/j.tics.2013.03.006](https://doi.org/10.1016/j.tics.2013.03.006)

Frankfurt HG (1978) The problem of action. *Am Philos Q* 15(2):157–162

Freeman WJ (2001) *How brains make up their minds*, 1st edn. Columbia University Press, New York

Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 9(10):474–480. doi:[10.1016/j.tics.2005.08.011](https://doi.org/10.1016/j.tics.2005.08.011)

Froese T, Di Paolo EA (2011) The enactive approach: theoretical sketches from cell to society. *Pragmat Cogn* 19(1):1–36. doi:[10.1075/pc.19.1.01fro](https://doi.org/10.1075/pc.19.1.01fro)

Froese T, Ikegami T, Virgo N (2012) The behavior-based hypercycle: from parasitic reaction to symbiotic behavior. In *Artificial Life*, vol 13, pp 457–464

Graybiel AM (2008) Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* 31:359–387

Hebb DO (1949) The organization of behavior: a neuropsychological theory, New edn. Psychology Press, Hillsdale

Heras-Escribano M, Noble J, de Pinedo M (2014) Enactivism, action and normativity: a Wittgensteinian analysis. *Adapt Behav*. doi:10.1177/1059712314557364

Hurley S (2001) Perception and action: alternative views. *Synthese* 129(1):3–40. doi:10.1023/A:1012643006930

Hutto DD (2005) Knowing what? Radical versus conservative enactivism. *Phenomenol Cogn Sci* 4(4):389–405. doi:10.1007/s11097-005-9001-z

Hutto DD, Myin E (2012) Radicalizing enactivism: basic minds without content. MIT Press, Cambridge

Iizuka H, Di Paolo EA (2007) Toward Spinozist robotics: exploring the minimal dynamics of behavioral preference. *Adapt Behav* 15(4):359–376. doi:10.1177/1059712307084687

Johnson MH (2001) Functional brain development in humans. *Nat Rev Neurosci* 2(7):475–483. doi:10.1038/35081509

Kelso JAS, Del Colle J, Schöner G (1990) Action-perception as a pattern formation process. *Attention and performance 13: motor representation and control*. Lawrence Erlbaum Associates Inc, Hillsdale, pp 139–169

Kohler I (1964) Formation and transformation of the perceptual world. International Universities Press, New York

Lenay C, Steiner P (2010) Beyond the internalism/externalism debate: the constitution of the space of perception. *Conscious Cogn* 19(4):938–952

Levy A, Bechtel W (2013) Abstraction and the organization of mechanisms. *Philos Sci* 80(2):241–261. doi:10.1086/670300

Llinas RR (2001) I of the vortex: from neurons to self. The MIT Press, Cambridge

Luhmann N (1986) The autopoiesis of social systems. In: Geyer F, van der Zouwen J (eds) *Sociocybernetic paradoxes*. Sage, London, pp. 172–192

Luhmann N (1995) Social systems. Stanford University Press, Redwood City

Luisi PL (2003) Autopoiesis: a review and a reappraisal. *Naturwissenschaften* 90(2):49–59. doi:10.1007/s00114-002-0389-9

Maturana HR (1970) Biology of cognition. In: Maturana HR, Varela F (eds) *Autopoiesis and cognition: the realization of the living*. Kluwer Academic, Dordrecht, pp 1–58

Maturana HR, Varela FJ (1980) *Autopoiesis and cognition*. D. Reidel Publishing Company, Dordrecht

Menary R (ed) (2006) Radical enactivism intentionality, phenomenology and narrative focus on the philosophy of Daniel D. Hutto. John Benjamins Publishing Company, Amsterdam/Philadelphia

Millikan RG (1989) In defense of proper functions. *Philos Sci* 56(2):288–302

Mingers J (1994) Self-producing systems: implications and applications of autopoiesis. Springer, Berlin

Moreno A, Etxeberria A, Umerez J (2008) The autonomy of biological individuals and artificial models. *Biosystems* 91(2):309–319. doi:10.1016/j.biosystems.2007.05.009

Moreno A, Ruiz-Mirazo K, Barandiaran XE (2011) The impact of the paradigm of complexity on the foundational frameworks of biology and cognitive science. In: Hooker CA, Gabbay DV, Thagard P, Woods J (eds) *Handbook of the philosophy of science*, vol *Philosophy of Complex Systems*. Elsevier, Amsterdam, pp 311–333

Mossio M, Taraborelli D (2008) Action-dependent perceptual invariants: from ecological to sensorimotor approaches. *Conscious Cogn* 17(4):1324–1340. doi:10.1016/j.concog.2007.12.003

Mossio M, Saborido C, Moreno A (2009) An organizational account of biological functions. *Br J Philos Sci* 60(4):813–841. doi:10.1093/bjps/axp036

Noë A (2004) Action in perception. The MIT Press, Cambridge

O'Regan JK, Noë A (2001) A sensorimotor account of vision and visual consciousness. *Behav Brain Sci* 24:939–1031

Piaget J (1969) Biology and knowledge. University of Chicago Press, Chicago

Piaget J (1975) *L'Équilibration des structures cognitives: Problème central du développement*. Presses universitaires de France

Piedrafita G, Montero F, Morán F, Cárdenas ML, Cornish-Bowden A (2010) A simple self-maintaining metabolic system: robustness, autocatalysis, bistability. *PLoS Comput Biol* 6(8):e1000872. doi:10.1371/journal.pcbi.1000872

Powers W (1973) Behavior: the control of perception, illustrated edn. Piscataway, Aldine Transaction

Prinz J (2006) Putting the brakes on enactive perception. *Psyche* 12(1):1–19

Ravaissón F (2008) Of habit. Continuum, London

Reinders AAT, Nijenhuis ER, Paans AM, Korf J, Willemsen AT, den Boer J (2003) One brain, two selves. *NeuroImage* 20(4):2119–2125. doi:10.1016/j.neuroimage.2003.08.021

Reinders AAT, Nijenhuis ERS, Quak J, Korf J, Haaksma J, Paans AMJ et al (2006) Psychobiological characteristics of dissociative identity disorder: a symptom provocation study. *Biol Psychiatry* 60(7):730–740

Rosslenbroich B (2009) The theory of increasing autonomy in evolution: a proposal for understanding macroevolutionary innovations. *Biol Philos* 24(5):623–644. doi:10.1007/s10539-009-9167-9

Rowlands M (2009a) Enactivism and the extended mind. *Topoi* 28(1):53–62. doi:10.1007/s11245-008-9046-z

Rowlands M (2009b) Extended cognition and the mark of the cognitive. *Philos Psychol* 22(1):1–19. doi:10.1080/09515080802703620

Ruiz-Mirazo K, Moreno A (2004) Basic autonomy as a fundamental step in the synthesis of life. *Artif Life* 10(3):235–259. doi:10.1162/1064546041255584

Ruiz-Mirazo K, Moreno A (2011) Autonomy in evolution: from minimal to complex life. *Synthese*. doi:10.1007/s11229-011-9874-z

Santos B, Barandiaran XE, Husbands P, Aguilera M, Bedia M (2012) Sensorimotor coordination and metastability in a situated HKB model. *Connect Sci* 24(4):143–161. doi:10.1080/09540091.2013.770821

Smith LB, Thelen E (2003) Development as a dynamic system. *Trends Cogn Sci* 7(8):343–348. doi:10.1016/S1364-6613(03)00156-6

Smithers T (1995) Are autonomous agents information processing systems? In: Steels L, Brooks R (eds) *The artificial life route to artificial intelligence: building embodied, situated agents*. Psychology Press, Hillsdale

Smithers T (1997) Autonomy in robots and other agents. *Brain Cogn* 34(1):88–106

Stewart JR, Gajenne O, Di Paolo EA (eds) (2010) *Enaction: toward a new paradigm for cognitive science*. MIT Press, Cambridge

Thompson E (2006) Sensorimotor subjectivity and the enactive approach to experience. *Phenomenol Cogn Sci* 4(4):407–427. doi:10.1007/s11097-005-9003-x

Thompson E (2007) Mind in life: biology, phenomenology and the sciences of mind, 1st edn. Harvard University Press, Cambridge

Thompson E (2011) Reply to commentaries. *J Conscious Stud* 18(5–6):5–6

Thompson E, Varela FJ (2001) Radical embodiment: neural dynamics and consciousness. *Trends Cogn Sci* 5(10):418–425. doi:10.1016/S1364-6613(00)01750-2

Tsuda I (2001) Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behav Brain Sci* 24(05):793–810

Tsukada M, Ichinose N, Aihara K, Ito H, Fujii H (1996) Dynamical cell assembly hypothesis—theoretical possibility of spatio-temporal coding in the cortex. *Neural Netw* 9(8):1303–1350

Varela FJ (1979) Principles of biological autonomy. North Holland, New York

Varela FJ (1992) Autopoiesis and a biology of intentionality. In: Proceedings of a workshop on Autopoiesis and Perception, pp 4–14

Varela FJ (1995) Resonant cell assemblies: a new approach to cognitive functions and neuronal synchrony. *Biol Res* 28(1):81–95

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

Varela FJ, Maturana HR, Uribe R (1974) Autopoiesis: the organization of living systems, its characterization and a model. *BioSystems* 5(4):187

Varela FJ, Thompson E, Rosch E (1991) The embodied mind: cognitive science and human experience. MIT Press, Cambridge, MA

Varela FJ, Lachaux JP, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. *Nat Rev Neurosci* 2(4):229–239

Villalobos M (2013) Enactive cognitive science: revisionism or revolution? *Adapt Behav* 21(3):159–167. doi:[10.1177/1059712313482953](https://doi.org/10.1177/1059712313482953)

Weber A, Varela FJ (2002) Life after Kant: natural purposes and the autopoietic foundations of biological individuality. *Phenomenol Cogn Sci* 1(2):97–125

Wolfram S (1984) Cellular automata as models of complexity. *Nature* 311:419–424

Wright L (1976) Teleological explanations. University of California Press, Berkeley