PhiMiSci
Philosophy and the Mind Sciences

Laying down a forking path:

Tensions between enaction and the free energy principle

Ezequiel A. Di Paolo[a,b,c] (ezequiel.dipaolo@ehu.es)
Evan Thompson[d] (evan.thompson@ubc.ca)
Randall D. Beer[e] (rdbeer@indiana.edu)

Abstract

Several authors have made claims about the compatibility between the Free Energy Principle (FEP) and theories of autopoiesis and enaction. Many see these theories as natural partners or as making similar statements about the nature of biological and cognitive systems. We critically examine these claims and identify a series of misreadings and misinterpretations of key enactive concepts. In particular, we notice a tendency to disregard the operational definition of autopoiesis and the distinction between a system's structure and its organization. Other misreadings concern the conflation of processes of self-distinction in operationally closed systems and Markov blankets. Deeper theoretical tensions underlie some of these misinterpretations. FEP assumes systems that reach a non-equilibrium steady state and are enveloped by a Markov blanket. We argue that these assumptions contradict the historicity of sense-making that is explicit in the enactive approach. Enactive concepts such as adaptivity and agency are defined in terms of the modulation of parameters and constraints of the agent-environment coupling, which entail the possibility of changes in variable and parameter sets, constraints, and in the dynamical laws affecting the system. This allows enaction to address the path-dependent diversity of human bodies and minds. We argue that these ideas are incompatible with the time invariance of non-equilibrium steady states assumed by the FEP. In addition, the enactive perspective foregrounds the enabling and constitutive roles played by the world in sense-making, agency, development. We argue that this view of transactional and constitutive relations between organisms and environments is a challenge to the FEP. Once we move beyond superficial similarities, identify misreadings, and examine the theoretical commitments of the two approaches, we reach the conclusion that far from being easily integrated, the FEP, as it stands formulated today, is in tension with the theories of autopoiesis and enaction.

Keywords
Agency · Autopoiesis · Enaction · Free energy principle · Sense-making

---

[a] Ikerbasque, Basque Foundation for Science, Bizkaia, Spain
[b] IAS-Research Center for Life, Mind and Society, University of the Basque Country, Donostia, Spain
[c] Centre for Computational Neuroscience and Robotics, University of Sussex, Brighton, UK
[d] Department of Philosophy, University of British Columbia, Canada
[e] Cognitive Science Program, Indiana University, Bloomington, USA

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

1 Introduction

The Free Energy Principle (FEP, Friston et al., 2006; Friston, 2012) has spurred a large body of literature in the sciences of mind. A growing subset of this literature concerns a series of claims about the compatibility between FEP and theories in embodied cognitive science, in particular, the enactive approach (e.g., Di Paolo et al., 2017; Thompson, 2007; Varela et al., 1991). Several articles (to date over two dozen) have appeared that make some connection between FEP, autopoiesis, and enaction. We are interested in examining these claims. They are varied and presented with justifications that range from pointing out general similarities to attempting to demonstrate deeper, principled connections between the two approaches. These claims have been interpreted in different ways. Some welcome the prospect of compatibility as positive to both the enactive and FEP positions, leading potentially to a more powerful and unified theory of embodied cognition. They speak of elaborations, interpretations, extensions, syntheses, etc. (e.g., Bruineberg et al., 2018). Others think FEP may serve to reconcile enaction with positions enactivists criticize, such as modern versions of representationalism (Clark, 2015; Constant et al., 2021; Wiese & Friston, 2021). Others adopt a different tone that suggests that FEP has overcome the limitations of the enactive approach, or subsumes or absorbs autopoietic and enactive theories (e.g., Allen & Friston, 2018; "FEP subsumes autopoiesis," Korbak, 2021, p. 2747; "FEP provides an implementation of enactivism, and in a sense supersedes or absorbs classical (i.e., autopoietic) formulations," Ramstead et al., 2021, p. 59). Finally, in contrast to the latter, some propose that enactive ideas can be used to fill in gaps and resolve problems in the FEP framework (Kirchhoff & Froese, 2017; van Es & Kirchhoff, 2021).

Given this variety, it is difficult to examine this literature as a coherent whole. Moreover, claims continue to evolve, so that sometimes important details absent in one piece of work become relevant in another. Nevertheless, we think that there are a few core shared claims that do not seem to change. In particular, there is the claim that the extension of FEP from its original domain of application in neuroscience to living systems in general serves as the basis of a broad theory of biology and cognition, something that resonates with enactive discourse about the continuity between life and mind (Thompson, 2007). A more specific common claim concerns the need to attend to the active role of the cognitive agent in engaging the world by following vital norms. This claim underlines the close connection between internal processes, processes in the environment, the agent's activity, and the agent's viability. At first sight, enactivists would agree with these claims. But are these parallels merely superficial? The literature we examine here claims that the connection between the two perspectives runs deep. We disagree. We will argue that the apparent compatibilities are based on quick readings, or even misreadings, of enactive ideas and what they entail, while central aspects of autopoiesis and enaction are left unacknowledged.

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

Clearing up misinterpretations may move the discussion forward, but will not by itself resolve the deeper tensions between FEP and enaction as theories of cognition. Many commentators focus on what they see as the main source of tension, which has to do with the representational and cognitive-loaded bias that instantiations of FEP and related ideas tend to adopt, particularly in the form of predictive processing models (Hohwy, 2013). They argue that FEP admits wider interpretations of terms such as "model" and "inference," and that these interpretations are compatible with enaction's rejection of internalism and computationalism. While these arguments have merit, we think they miss much more basic tensions between the two theories. These tensions have to do with how the enactive approach conceives of agents as precarious, self-constituted entities in ongoing historical development and capable of incorporating different sources of normativity throughout their development, a world-involving process that is co-defined with their environment across multiple spatiotemporal scales and together with other agents. The enactive view, as we shall explain, is at odds with universalizing the ideas of non-equilibrium steady states and Markov blankets that serve as key assumptions in FEP. Underlying these worries are divergent fundamental conceptions of materiality and temporality. These incompatibilities lead to a series of forking claims about embodied agents in general, claims that take on a sharper contrast in the case of human beings due their diversity and historicity. Nonrepresentational interpretations of FEP do not rejoin these divergences.

Before we proceed, a few caveats. First, we do not intend to critically evaluate FEP altogether (for this, see e.g., Aguilera et al., 2021; Baltieri et al., 2020; Biehl et al., 2021; Colombo & Wright, 2021; Litwin & Miłkowski, 2020; Raja et al., 2021; van Es, 2021). We are exclusively concerned with examining claims about the compatibility between FEP and enaction. Second, in specifying what we see as points of theoretical tension between the approaches, we work on the basis of the FEP as currently formulated. We make no claims that clarifications, extensions, or elaborations of the FEP could not address some of the divergences between the approaches.

Our argument will proceed roughly as follows: (1) There are certain principles and phenomena that FEP needs to be able to account for if it claims to make contact with enaction; (2) to date, FEP has not successfully accounted for these enactive principles and phenomena, despite some claims to the contrary in the literature; (3) there are reasons to think that it might be difficult for anything like the FEP to provide such an account, although time will tell and we are certainly not claiming any impossibility proof. Finally, we do not reject specific applications of FEP ideas or techniques to make sense of empirical data (Walsh et al., 2020). Insofar as they offer useful methods for analysis, FEP and predictive processing models can provide helpful tools in specific cases. Our discussion concerns FEP as a *general* approach to explaining life and mind. It is at this level that we find significant and hard to reconcile differences with the enactive approach.

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

2 An enactive FEP?

Enactive readings of FEP and claims of compatibility with ideas such as autopoiesis are varied. Some attempt principled unifying arguments (Ramstead et al., 2021), others provide readings of FEP from enactive and ecological perspectives (Bruineberg et al., 2018), and others simply remark on apparent resonances and similarities (Clark, 2015).

Despite this variety, the motivations for these diverse claims seem fairly aligned. We can roughly describe them as a concern to establish embodied-cognition interpretations of FEP that do not necessarily entail a traditionally internalist or representationalist perspective (as predictive processing models tend to do, Hohwy, 2013), or that at least recast notions of representation in action-oriented terms (Clark, 2015). Some authors are also motivated by the potential extensions of FEP, originally formulated within the confines of neuroscience, to a wider range of biological phenomena—indeed, to the status of a general principle of theoretical biology and psychology. Demonstrating compatibility with the enactive approach, with its emphasis on the continuity between life and mind, could be considered as a step towards this goal.

Because of these aligned motivations, it makes sense to look at this literature as a whole. We group its claims into two broad sets. We are aware that we may miss some nuances, but we do not think these will be central to our purposes. In particular, we are concerned with the following two groups of ideas: (1) proposals for a link between FEP and the theory of autopoiesis, and (2) discussions of various compatibilities between FEP and enactive ideas.

We present only selected examples in this section and offer a critical analysis in the next two sections. We will not occupy ourselves with other discussions that point to broad similarities between FEP and "enactive-like" ideas (e.g., situated embodied activity, action-perception co-dependence) that are shared by various other approaches (e.g., Pezzulo et al., 2015).

It is important to keep in mind that autopoiesis and enaction are obviously related and the division into two groups is merely one of convenience. The unspoken assumption is that the enactive approach has developed historically from readings and elaborations of the ideas first presented in the classical theory of autopoiesis (Maturana & Varela, 1980, 1987), and that to this day many central concepts in enaction find their roots in this theory. However—and this may be a confounding factor to keep in mind—the enactive approach is in many ways a branching development of the theory of autopoiesis and is critical of important aspects of classical autopoietic theory (e.g., Di Paolo, 2018). Contemporary defenders of classical autopoietic theory, in turn, have criticized the enactive approach precisely for this reason (e.g., Villalobos & Ward, 2015).

Our first group concerns the goal championed by Karl Friston and others of extending the reach of the FEP to biology. This work discusses FEP's relation to the theory of autopoiesis and concerns enactive ideas only indirectly. Nevertheless, its

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

claims form a kind of backbone for work looking at enaction, the reasoning being that enactive theory is an elaboration of the theory of autopoiesis, as we have just said.

Friston (2013) provides an early claim concerning the relation between FEP and autopoiesis (for more up-to-date versions of this claim, see Parr et al., 2019; Ramstead et al., 2021; Wiese & Friston, 2021). This paper describes the applicability of FEP to biological systems (see also Friston, 2012) by explicitly referring to the concept of *autopoiesis*.

Briefly, autopoiesis is a property of the organization of living systems (Maturana & Varela, 1980, 1987). Maturana & Varela (1980) originally postulated it as the defining property, necessary and sufficient for the living organization, but subsequent researchers have argued that autopoiesis is necessary but not sufficient for the living organization (Bitbol & Luisi, 2004; Bourgine & Stewart, 2004; Di Paolo, 2005; see Thompson, 2007, for discussion). "Autopoiesis" means that living systems are organized as networks of biochemical processes such that two conditions obtain: (1) (self-production) the operation of the processes in the network regenerates the set of relations between processes in the network; and (2) (self-distinction) the network emerges as a distinct topological unity in the domain of biochemical interactions. An autopoietic system is therefore materially self-producing and self-distinguishing. Due to the circularity (recursive self-production) to which this idea refers, Maturana & Varela (1980, see also 1987; Varela, 1979) described an autopoietic system as manifesting "organizational closure" or "operational closure." We will come back to these ideas in the next section.

Friston (2013) considers an ergodic random dynamical system. According to him, ergodicity in this context "means that the time average of any measurable function of the system converges (almost surely) over a sufficient amount of time" (ibid., p.2) and that "one can interpret the average amount of time a state is occupied as the probability of the system being in that state when observed at random" (ibid.). Such a system will tend to evolve towards a regime where the probability density of its states (the chance that a particular state will be visited) does not change over time.

If the density of states is unchanging, that is, if the system reaches a non-equilibrium steady state (NESS), one can derive a series of relations between internal and external variables. The distinction between these sets of variables is defined by assuming the presence of a Markov blanket that statistically "insulates" one set from directly affecting the other.[1] Instead, effects propagate through action and sensory variables that constitute the blanket and mediate between internal and

---

[1] Biehl et al. (2021) clarify that Markov blankets are presented as a combination of two independent conditions. One concerns the vector field underlying the stochastic differential equation of the system forcing any systematic influences between external and internal variables through the sensory and active variables. The other is an actual Markov blanket condition on the stationary distribution saying that according to this distribution the internal and external states are conditionally independent given sensory and active states. For our purposes we take Markov blankets to imply a combination of the two conditions.

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

external variables. According to Friston, these relations demonstrate the central claim that "any ergodic random dynamical system that possesses a Markov blanket will appear to actively maintain its structural and dynamical integrity" (Friston, 2013, p. 2, emphasis removed). Friston sees the maintenance of structural and dynamical integrity as the key property of biological systems, from which he conjectures that, to exist, organisms must be ergodic[2] (or converge to a NESS, e.g., Costa et al., 2021; Parr et al., 2019) and must be in possession of a Markov blanket.[3]

Friston refers to autopoiesis as an attribute of biological self-organization and claims that it can be derived from his analysis. He sees the maintenance of a steady state as a kind of structural integrity; and he sees such structural integrity as a form of homeostasis sustained by the system appearing to act against the dispersion or entropy of its states (Friston, 2013, p. 5). The same argument is repeated in more recent publications (e.g., Parr et al., 2019; Wiese & Friston, 2021). The parallel is reaffirmed elsewhere by suggesting that Markov blankets play the role of the autopoietic system's boundary-subserving processes of self-distinction. Allen & Friston (2018) further claim that "One can formulate this in another way; the organism's internal states constitute probabilistic beliefs about what actions are the most likely to provide evidence for the organism's existence (survival)" (ibid. p. 2474), and they state that "This notion is at the heart of autopoietic views of life and mind, insofar as it induces a deeply circular causality between internal and external states, to provide a normative principle by which to understand all action and perception" (ibid. p. 2474).

---

[2] The phrase "biological systems are ergodic" found in (Friston, 2013, p. 5) could be interpreted as involving both the organism and its environment. Similarly with the phrase: that "all living systems revisit a bounded set of states repeatedly (i.e., they are locally ergodic)" (Ramstead et al., 2018, p. 3) although from the context "living systems" seems equivalent to "organisms." Friston (2012, p. 2105), however, makes reference to systems "in a changing (and possibly non-ergodic) environment" suggesting that the ergodicity claim is made about organisms.

[3] Friston's argument bears a resemblance to an idea presented by W. R. Ashby (1962) if we interpret his use of the term "equilibrium" to mean a steady-state flow. "So the answer to the question: How can we generate intelligence synthetically? is as follows. Take a dynamic system whose laws are unchanging and single-valued, and whose size is so large that after it has gone to an equilibrium that involves only a small fraction of its total states, this small fraction is still large enough to allow room for a good deal of change and behavior. Let it go on for a long enough time to get to such an equilibrium. Then examine the equilibrium in detail. You will find that the states or forms now in being are peculiarly able to survive against the changes induced by the laws. Split the equilibrium in two, call one part 'organism' and the other part 'environment': you will find that this 'organism' is peculiarly able to survive against the disturbances from this 'environment.' " (ibid. p. 272). This resemblance may explain Friston's insistence on using the language of "appearances," e.g., internal states appearing to minimize energy, appearing to engage in active inference, and appearing to model the external world, to mention three instances appearing in the abstract in Friston (2013). Ashby's exercise, like Friston's, is ambiguous between deflating the notion that there is a real distinction between organism and environment, when in fact the boundary is drawn by the observer (or the choice of one of possibly many Markov blankets), and suggesting that to speak of actual distinctions and of their appearances is ultimately the same thing, a view that enactivists tend to reject (e.g., Barandiaran et al., 2009; Di Paolo et al., 2017, p. 39).

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

The central claim of these arguments is that random dynamical systems with one or several Markov blankets that converge to a NESS conserve themselves in that stationary condition (or appear to do so depending on the perspective) by minimizing free energy, and that these conditions are equivalent to the condition of a system being autopoietic.

As we discuss later, Friston's (2013) use of the concept of autopoiesis is rather loose. He does not attempt to substantiate his claim by considering the definition of autopoiesis; instead, he bases the claim on a broad analogy between the conservation of systemic properties in the random dynamical systems he proposes and in autopoietic systems. Further elaborations of his argument move into more specific enactive territory and lean more explicitly on ideas such as *operational closure*, *adaptivity*, *agency*, and *sense-making* (Di Paolo, 2005, 2009; Thompson, 2007). For instance, Kirchhoff et al. (2018) discuss the significance of the concept of the Markov blanket. After pointing out that metabolic cell production is a process enabled by the existence of cellular boundaries, the authors conclude that "living systems can therefore be construed as a process of boundary conservation, where the boundary of a system is its Markov blanket" (ibid. p. 6). The dependencies induced by the Markov blanket act as a "kinetic barrier" that keeps the system "far removed from thermodynamical equilibrium." A Markov blanket does not fully isolate the system: "external states may influence internal states even if the former are not constitutive parts of an operationally closed system" (ibid). Ramstead et al. (2021) draw similar analogies between Markov blankets and the boundaries of (unicellular) organisms. More specifically, they claim in agreement with Kirchhoff et al. that "it is fairly straightforward to establish that the Markov blanket formalism provides a statistical formulation of operational closure" (Ramstead et al., 2021, p. 55).

Our second group in the literature we are discussing corresponds to arguments in favour of a broad spectrum compatibility between FEP and enaction, not just the concept of autopoiesis, though in all cases the claims tend to rely on the work in the first group. For instance, Bruineberg et al. (2018) accept that FEP is applicable to biological systems, but they reject its association with Helmholtzian approaches, in which perception is mediated by unconscious inferences about the causes of sensation. A typical version of such approaches is the postulation of predictive processing models implemented in the brain. Bruineberg and colleagues argue that FEP is more general and one need not be committed to a Helmholtzian perspective (though even when conceiving of strictly biological integrity in terms of the FEP, Friston already talks about active inference or the semblance of it). The authors articulate an ecological-enactive interpretation of FEP, avoiding the Helmholtzian sharp separation between organism and environment, and placing the locus of active inference in the whole situated organism and not just its brain. They offer charitable interpretations, such as construing the internal dynamics of active inference as states of action-readiness or expounding Friston's claim that an organism embodies an optimal model of its ecological niche (Friston, 2011) in terms of the organism possessing adequate skills that allow it to reach conditions of grip.[4] These interpretations are indeed more palatable to enactivists and ecological psychologists than the Helmholtzian picture.

Applying a similar ecological-enactive interpretation of FEP to questions concerning the self and the feeling of mineness, Kiverstein (2020) offers a more detailed comparison with enactive technical concepts. Like Ramstead et al. (2021), he describes the concept of autonomy in terms of precarious operational closure and suggests that this concept can be cashed out in FEP terms. According "to the free energy principle, the autonomy of living systems is a consequence of the inferential processes of free energy minimisation" (Kiverstein, 2020, p. 565), and "the free energy principle tells us how living systems might sustain their own operational closure under precarious conditions in their dynamic coupling with the environment" (ibid. p. 566).

Kiverstein also makes reference to key aspects of the enactive concepts of agency and sense-making:

> Any system that has autonomy will also qualify as an agent that has its own individual point of view upon the world (Di Paolo et al., 2017). Relative to this point of view the environment has affective significance in terms of how it bears on the organism's self-produced identity. Organisms enact values, purposes and norms which are of their own making in the sense that they originate in processes of self-individuation (i.e. free energy minimisation) to which the organism owes its existence. […] Perception and action are thus laden with affect […]. (Kiverstein, 2020, p. 566)

Except for the parenthetical reminder that he accepts the equating of self-individuation and free energy minimization, Kiverstein in these passages does not seem to be arguing for a parallel between enaction and FEP as much as for what an enactive interpretation of sense-making could add to FEP.

Kiverstein also suggests that a more embodied and situated understanding of active inference should be in terms of readiness for action, and he uses recent developments in enactive theory to expand on what this idea means.

> As the organisation of the autonomous system becomes less bound to its immediate metabolic needs, so the possible meaningful relations the organism can stand in to the environment becomes less tightly bound to the here and now. The organism becomes sensitive to tendencies and trajectories that constitute the dynamical configurations of the organism–environment system, and their consequences for its precarious existence […] The nervous system can then be thought of as generating and sustaining stable and recurrent patterns of sensorimotor engagement with the environment. These patterns of engagement

---

[4] Recent interpretations of Friston's claim de-emphasize the idea that the organism implements or is itself a model of its environment, and argue instead that these models correspond to descriptions available to the observer (see, e.g., Baltieri et al., 2020; van Es, 2021).

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

> with the environment exhibit just the same properties of operational closure and precariousness as we find in the more basic processes of homeostasis. The argument [here] is that patterns of sensorimotor engagement owe their operational closure to processes of free energy minimisation. (Kiverstein, 2020, p. 566)

In contrast to Kiverstein's attempt at a careful inter-theoretical comparison, other commentators offer their own particular reading of enactive ideas from an FEP standpoint. For example, Clark (2015) defends an embodied and action-oriented construal of predictive processing, which he uses to interpret Varela et al.'s (1991) notion of "enacting a world." Despite there being several elaborations of this notion in the enactive literature (Stewart et al., 2010; Thompson, 2007; Varela, 1991a, and others), Clark idiosyncratically suggests that one way to interpret phrases like "bringing forth a world" is through processes such as active data sampling used to sustain and update a cognitive agent's predictive models in an action-dependent manner.

Kirchhoff (2018) follows the thrust of the argument by Friston (2013), adding a few glosses: "FEP adds more to life than AT [autopoietic theory] [...] for the FEP, living systems minimize surprise on the basis of embodying a probabilistic model of themselves and their environment. This is a step beyond the appeal to mere AT in explaining the process of self-maintenance" (Kirchhoff, 2018, p. 2526). Following enactive arguments about the insufficiency of bare autopoiesis for sense-making, Kirchhoff states that the FEP "incorporates *adaptivity*—viz., the future-oriented aspect of cognition—from the very beginning." This point justifies the perspective that, more than just being compatible with autopoiesis, FEP is also compatible with the enactive approach. Adaptivity (Di Paolo, 2005) is taken as a complexification of autopoiesis. "This shoehorns nicely with the FEP. Given that the FEP can be shown to apply to systems that are, at least arguably, not cognitive, it follows that mentality arises when organisms minimize free energy to a certain degree—viz., in the context of active inference" (ibid. p. 2535).

Kirchhoff continues, however, by raising the potential worry that FEP might not fully account for the enactive notion of sense-making and the existential-phenomenological sense of life-mind continuity: "nothing in the FEP is able to account for the constitution of a meaningful perspective" (Kirchhoff, 2018, p. 2535). After describing the idea of sense-making as the most general aspect of mindedness and acknowledging that for enactivists this idea is central to an explanation of life-mind continuity, Kirchhoff asks whether this concept can fully be captured by the minimization of free energy. This is one of the few instances in the literature we are examining where a potential incompatibility (rather than a mere difference) between FEP and enaction is explicitly signaled: "if the FEP gives up on sense-making, then how can it explain what is central to life and mind?" (ibid. p. 2536). Kirchhoff observes that FEP does not negate the idea of sense-making (so the relation between them is not so much compatibility as non-incompatibility), and that this idea has itself been criticized by so-called radical enactivists (Hutto & Myin, 2013) for possibly entailing that basic cognition has meaningful content (content subject to correctness conditions), contrary to the nonrepresentationalist aims of the enactive approach (but see Thompson, 2018, for a critical response to radical enactivism). In a strange reversal of roles, Kirchhoff argues that if there were such an implication leading from sense-making to content, then the concept of sense-making, and hence the enactive approach, would be in tension with the nonrepresentational, leaner version of FEP he defends. Here an enactive explanation of meaning moves from something that FEP as a theory of the mind lacks to something it should not have because it seemingly invites representations back.

Elsewhere we see similar worries about FEP's potential lack of cognitive specificity, as in Kirchhoff & Froese (2017), who suggest that enactive theory can constrain overly broad interpretations of FEP. As we mentioned, Friston's (2013) argument applies to any random dynamical system in a NESS with a Markov blanket, and so the conclusions should be valid for a wide range of systems, including systems traditionally considered non-living or non-cognitive. Kirchhoff and Froese argue that Friston should therefore be committed to a view that sees mind everywhere, as part of any system that minimizes free-energy, including non-biological ones (or, given some kind of life-mind continuity, to a view that finds both life and mind nearly everywhere). Given the existence of systems, such as a candle flame, or in Bruineberg et al.'s (2018) example, two synchronized pendulum clocks, whose stable features are more parsimoniously explained through dissipative self-organization or dynamical couplings than by supposing that internal variables are performing active inferences, Kirchhoff and Froese are right to worry about a thinning out of the explanatory power of the FEP. They suggest that life and mind should be seen as strongly continuous, and that this continuity places cognitive specificity demands that FEP does not meet. To remedy this problem, they suggest that enactive versions of life-mind continuity, defined in terms of autonomy, adaptivity, and sense-making, should be used to constrain an otherwise too liberal FEP.

In this quick exposition we have seen a sample of how FEP, autopoiesis, and enaction have been discussed almost exclusively in terms of important compatibilities. These compatibilities are of different kinds, ranging from broad analogies between free energy minimization, Markov blankets, and aspects of autopoiesis to more elaborate readings using enactive concepts such as autonomy, adaptivity, and sense-making. These compatibilities serve to indicate the possibility of a more detailed inter-theoretical relation or to supplement what some authors see as problematic aspects of FEP as a candidate universal approach to life and mind. We find very few instances where serious incompatibilities are mentioned, making things look as if, except for some details and clarifications, there is a broad consensus about the match between the approaches.

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

Before we discuss why we strongly disagree with this picture and we examine deeper problems that have not been raised by any of these authors, we will mention a few misreadings that we consider important to clarify for this and future discussions.

3 Misreadings of autopoiesis and enaction

The enactive approach owes much to the classical theory of autopoiesis but it is not identical with it. Enaction inherits several concepts from this theory, and more importantly, a sensitivity to the need to provide operational definitions of its key ideas. As well as serving as a constraint on theory development, such definitions should also facilitate inter-theoretical debates. Concepts such as autopoiesis, autonomy, operational closure, sense-making, and agency are defined in operational terms in the enactive literature to avoid vague interpretations. It is surprising that very few attempts to link FEP, autopoiesis, and enaction have taken advantage of these operational definitions and that technical terminology is sometimes used loosely or interpreted narrowly. This sloppiness results in a series of misreadings that, as we argue later, can signal deeper incompatibilities between the approaches. Here we indicate some of these misreadings.

3.1 Autopoiesis

Let us start by re-stating the classical definition of autopoiesis (after Maturana & Varela, 1980, pp. 78–79):

> An autopoietic system is organized (defined as a unity) as a network of processes of production (transformation and destruction) of components which:
> (i) through their interactions and transformations continuously regenerate and realize the network of processes and relations that produces them; and
> (ii) constitute the system as a concrete unity in the space in which the processes exist by specifying the topological domain of its realization as a network.

We call the first and second requirements "self-production" and "self-distinction," respectively. It is possible to show that both requirements are dialectically related to each other by using the concept of *adaptivity* (Di Paolo, 2018), which is important in recent enactive literature (Di Paolo & Thompson, 2014; Thompson, 2007). We return to this concept later.

Friston (2013, p. 5) refers to autopoiesis as an attribute of biological self-organization and homeostasis in which "active states will appear to maintain the structural and functional integrity of biological states." Others make similar remarks, such as "the internal and blanket states that constitute a subsystem are

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

autopoietic, because their (nonequilibrium steady-state or ergodic) probability density is maintained by the flow of the subsystem's internal and active states." (Palacios et al., 2020, p. 5). Autopoiesis is here understood as sufficiently characterized by self-organization and integrity, entailing persistence over time.

However, autopoiesis is not *defined* as self-organization, integrity, or persistence over time. The definition describes the organization of a system that jointly fulfils the requirements of self-production and self-distinction. A system may self-organize and not be autopoietic. It may persist over time and not be autopoietic. And it may spontaneously achieve integrity and still not be autopoietic. The dynamic processes and conditions that form a crystal in a supersaturated liquid solution result in integrity, self-organization, and persistence over time but they do not constitute an autopoietic system (it does not continuously regenerate and realize the network of processes and relations that produces the crystal; it simply undergoes a one-off individuation event until it reaches a steady state). Autopoiesis may in many cases entail integrity, self-organization, and persistence over time. When present, these properties are realized in specific ways in autopoietic systems, i.e., by fulfilling the requirements of self-production and self-distinction. But the presence of these properties in a given system, as we see in the crystal counter-example, does not, in any combination, entail autopoiesis. The literature linking FEP and autopoiesis often takes these attributes as sufficient and in doing so disregards important aspects of the definition.

To clarify these aspects, classical autopoietic theory introduces a distinction between the structure and the organization of a system (Maturana & Varela, 1980). The structure is the system's actual realization, the concrete components that constitute a system and the actual and concrete relations between them. The system's organization is the abstract set of relations that define the system as belonging to a class. Autopoiesis is the description of a class of systems, i.e., a description of the organization that defines this class. Concrete autopoietic systems may be instantiated in a wide variety of structures, and a given structure may belong to more than one class of organization (Fido is a dog, a mammal, a living organism). Structures also change over time, even if the organization remains invariant (Fido was a puppy, is an adult hound, will be a lazy senior). The distinction between structure and organization, as well as other technical concepts, such as *structural coupling* and *operational closure*, are further clarified by Beer (2015; Beer, 2020) using a toy model that reveals other subtleties about these ideas, such as the relevance of their spatiality and time-extendedness.

To come back to the relation between autopoiesis and integrity, self-organization, and persistence over time, it should now be clear that we need to be careful about whether we are talking about structure or organization. For instance, the autopoietic organization of a living organism persists over time as long as it lives, but its structure most often does not. Statements about the structure of a system do not obviously translate into statements about its organization, and vice versa.

Setting aside this misreading of the idea of autopoiesis we may still ask whether the formal proposal made by Friston and colleagues can be expected to lead to systems that specifically meet the requirements of self-production and self-distinction. We address this question in more detail in Section 4, but here we discuss why Friston and colleagues might think that this is the case.

Autopoiesis entails the conservation of biological organization and an adaptive relation to the environment, a relation that allows the autopoietic system to remain viable. Conservation of organization has sometimes been illustrated in the classical autopoietic literature as a special form of homeostasis, lending some credence to Friston's interpretation of biological systems as resisting a tendency to disorder (e.g., Friston, 2011, p. 92). For instance, "an autopoietic machine is an homeostatic (or rather a relations-static) system which has its own organization (defining network of relations) as the fundamental variable which it maintains constant" (Maturana & Varela, 1980, p. 79). The qualification ("relation-static") should be read as a warning that we are not talking here of homeostasis in the traditional dynamical sense. Indeed, the idea of an autopoietic *organization* as a regulated variable is problematic and not generally accepted. At the organizational—as opposed to the structural—level there is no gradient to the condition of being or not being autopoietic that could be thus regulated (Di Paolo, 2005).

The sense of homeostasis to which Friston and colleagues refer correspond to structural properties not organizational ones. While several references to "structural integrity" and to "self-organization" can be found in Friston (2013) and other work, none of them corresponds to the structure/organization distinction elaborated in autopoietic theory. The structural integrity of systems with a Markov blanket as they converge to a NESS bears some resemblance to the idea of a system that actively conserves its organization. But it is not the same idea.

A contrast with the definition of autopoiesis can verify this claim. Several discrepancies become apparent. There is no obvious organization-level equivalent in Friston's systems to the network of processes that through transformations *realize* the conditions of its own production. In fact, the assumptions (of an invariant density of states and the presence of a Markov blanket) are such that no regeneration is needed after the NESS is reached; the "organization" of the system simply endures. It is unclear in what sense the components of the systems, e.g., its Markov blanket, are materially *produced* by other processes in the system, instead of just being there by assumption (see also Raja et al., 2021 for a similar point). Friston himself may be thinking of this problem when he recognizes that more than entropy minimization is needed to distinguish a biological system from a petrified stone (Friston, 2013, p. 11).

In summary, autopoiesis has been loosely interpreted and indeed often misread to different degrees by Friston and colleagues. Attending to the operational definition of autopoiesis and the distinction between structure and organization, it is possible to show that the characterization of autopoiesis presented by Friston and colleagues (structural integrity, persistence, self-organization) does not entail autopoietic organization; such properties can be found in non-autopoietic systems. Nor do autopoietic systems necessarily realize such properties in their structure. Lack of attention to the structure/organization distinction can lead to readings of autopoiesis where both are conflated.

3.2 Boundaries

The concept of operational closure under precarious conditions (explicated in Di Paolo & Thompson, 2014) underpins the enactive idea of *autonomy* (derived from Varela, 1979), another technical term that extends the concept of autopoiesis to more general domains:

> [An] autonomous system is defined as a system composed of several processes that actively generate and sustain an identity under precarious circumstances. In this context, to generate an identity is to possess the property of operational closure. This is the property that among the conditions affecting the operation of any constituent process in the system there will always be one or more processes that also belong to the system. And, in addition, every process in the system is a condition for at least one other constituent process, thus forming a network. (Di Paolo, 2009, p. 15)

In other words, the processes that constitute an operationally closed network relate to each other such that they form a set of mutually enabling relations. They do so under precarious conditions, meaning that in the absence of these mutually enabling relations, the same processes would tend to run down. No component is, in other words, strictly self-standing in the absence of the whole network.

Operational closure does not mean that processes external to the network cannot influence those that belong to it, or that processes in the network cannot influence processes external to it. Nor does it mean that these influences cannot have an enabling character, i.e., by being causally necessary for internal processes to continue to exist. In other words, an operationally closed system is open not only to informational external influences but also to all kinds of material and formative relations, including, in the case of living systems, the transport of heat, mass, biomatter, genetic material, microorganisms, and so on. Internal processes can literally depend on such enabling exchanges to continue to operate.

Ramstead et al. (2021) and Kirchhoff (2018) make reference to operational closure. These authors see in this concept an enactive analogue of a Markov blanket. For Ramstead et al. (2021, p. 555), as we have indicated earlier, “it is fairly straightforward to establish that the Markov blanket formalism provides a statistical formulation of operational closure.” For Kirchhoff (2018), FEP accommodates, through the idea of Markov blankets, the requirement of operational closure and goes beyond it. Confusingly, Kirchhoff, following Friston (2013), also states that it “can be shown that Markov blankets operate in much the same way as a cell boundary” (Kirchhoff, 2018, p. 2527; see also Allen & Friston, 2018, p. 2473; Kirch-

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

hoff et al., 2018, p. 6). We also read that a “cell therefore has a Markov blanket—its plasmalemma [cell membrane]” (Ramstead et al., 2021, p. 551). More complex organisms are described as bounded by an ensemble of nested Markov blankets (Kirchhoff et al., 2018). In addition, Markov blankets are said to act as the epistemological and ontological boundaries of cognitive, not just biological, systems (Ramstead et al., 2021, p. 551). Similarly, Kirchhoff & Kiverstein (2021) propose using the formalism of Markov blankets to delineate flexible boundaries for the mind.

Such statements are speculative and perplexing. They again confuse the organization and the structure of living and cognitive systems. We see no straightforward relation between operational closure (an organizational statement about how a network of processes actively produces and distinguishes itself) and Markov blankets (a statement about statistical conditional independence between sets of variables). Very few commentators seem to have remarked on this discrepancy (e.g., Bruineberg & Hesp, 2018, p. 38; see also Bruineberg et al., 2021). As we have said, an operationally closed system is open to all kinds of interactions and exchanges with the environment, as long as its organization is not destroyed. These exchanges are typically regulated by the system, but not always (think of exposure to ionizing radiation or the effects of gravity), and not always successfully (think of a viral infection). Potential breakdowns caused by unregulated exchanges are not necessarily fatal, meaning the condition of operational closure can still be sustained. Because operationally closed systems may be realized in such complex unmediated relations between internal and external variables, it seems unreasonable to assume that operational closure is therefore equivalent to the statistical conditions imposed by Markov blankets.

Moreover, because operational closure is an organizational concept, its relation to biological or cognitive *structural* “boundaries” is not obvious nor are there obvious relations between the boundaries of biological and cognitive systems, as discussed in Di Paolo (2009). If we know how the organization of an operationally closed system is structurally instantiated, we could in principle point to processes that belong to it and processes that do not. But this epistemic operation is not replaceable by the act of pointing to a physical structure, such as a cell membrane. Maintaining the condition of self-distinction can lead to processes that regulate complex, structural and spatial boundaries, as in the case of cell membranes, but also to other processes, such as the prokaryotic CRISPR-Cas immune system that defends unicellular organisms from viruses (Rath et al., 2015). Importantly, these self-distinction processes, including membranes, are not in any organizational sense at the boundary of the operationally closed system; rather, they are *part of it*, they are *in it*.

In sum, the status of the relations between different concepts, such as the systemic distinction enabled by operational closure, the structural processes that subserve self-distinction, the presence of actual spatial boundaries, and the conditional statistical independence between internal and external variables (Markov blankets) is anything but “fairly straightforward” (Ramstead et al., 2021, p. 555).

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

4 Points of tension between the two approaches

Attempts to establish some sort of connection between FEP and enaction continue to appear as we write these lines. Some claims change in tone and detail, but so far we have not seen many cases in which the misreadings we have indicated are not present in some form or other. In this section, we raise points of deeper tension between the two perspectives. As stated in the Introduction, our argument is that there are certain principles and phenomena that FEP needs to account for if it claims to make contact with or encompass enaction, and that, to date, FEP has not successfully accounted for these enactive principles and phenomena. There are reasons to think that it might be difficult for future developments of anything like FEP to account for these enactive principles and phenomena, although we are not claiming any impossibility proof.

Let us clarify some terminology. The comparison between different theoretical approaches can be contentious precisely because it is difficult to agree on a common ground about what aspects should be contrasted. These aspects can include explicit and implicit assumptions, ontologies, types of causality, semantics, categorizations, standards, methodologies, formalizations, models, interpretations, explanatory goals, intended scope, and applications. The criteria for declaring two theories to be incompatible vary and can be contentious if we are not clear about exactly what we are comparing. To facilitate the reading of this section we will use the term *tension* to describe an apparent contradiction between theories. A tension might turn out to be a *challenge*, i.e., a contradiction that might be resolved by further work and conceptual coordination not involving a change to fundamental assumptions (e.g., corrections, extensions, clarifications, or making explicit something that was only implicit up to this point). We will qualify tensions as *incompatibilities* when, within the scope of our analysis, statements describing assumptions, categories, formalizations, and/or interpretations in each theory cannot be held jointly together. Claims of incompatibility in this section can be contested by pointing to limitations of the analysis or by offering reformulations, possibly turning them into challenges. However, the point is that such reformulations would require major theoretical accommodation, such as abandoning basic assumptions. Because of this, to show an incompatibility is not the same as presenting an impossibility proof, since the latter requires shared terminology and frameworks that may not be currently in place, as in the case here.

A stronger difference between theories is that of *incommensurability* (Kuhn, 1962). This refers to the difficulty of comparing the holistic nature of scientific theories (involving in addition to assumptions, formalizations, interpretations, standards, etc., also technological, institutional, economic, and cultural aspects). We do not attempt to claim incommensurability between FEP and autopoiesis/enaction, but we do note that it is sometimes possible to appreciate a divergence in the style and orientation of discourses between these approaches. These lead to a difficulty in "neutrally" comparing assumptions, semantics, and formal aspects because these hold different value to different researchers, a difficulty we acknowledge in our case.

Our task is not helped by the ongoing revisions and refinements to the FEP. To be as fair as possible, we will attend to the more recently published versions and variants of these claims.

4.1 Non-equilibrium steady states versus history

The work presented in Friston (2013) is central to practically all the proposals comparing FEP and enaction. It continues to be cited approvingly to this day. It has been argued that the formal argument presented in this work has technical problems (Biehl et al., 2021; see also Friston et al., 2021), and that when applied to concrete systems "the assumptions of the FEP do not hold for a very broad class of systems, namely linear, weakly coupled systems, except for the limited case of fully symmetric agent-environment interaction" (Aguilera et al., 2021, p. 18). We will not focus on these problems. Our goal is to examine claims of compatibility with the enactive approach. To do this, it turns out that it is sufficient to examine the central assumptions of the FEP because we already find tensions between the two theories at this starting point. Here we question whether the convergence to non-equilibrium stationary states, or NESS, which has been and remains central to the FEP argument, is compatible with the theoretical claims of the enactive approach.

The assumption we are concerned with is that the random dynamical system in question will converge to an invariant set of states: "Because the system is ergodic (and weakly mixing) it will, after a sufficient amount of time, converge to an invariant set of states called a *pullback* or *random global attractor*" (Friston, 2019, p. 10). In other words, the system reaches a NESS with an invariant density of states (Costa et al., 2021; Friston, 2019; Palacios et al., 2020; Parr et al., 2019). This enables an important step in the formal derivation of FEP mathematical relations. The invariant NESS density describes the probability of visiting particular states even if the variables continue to change. As a consequence, the flow of the system can be expressed as a time-independent operation on this invariant probability; see e.g., equation 2.3 in Friston (2013), equations 3.2 in Parr et al. (2019), equations 1.8 in Friston (2019), equations 2 in Palacios et al. (2020). The time invariance of these expressions is key: setting "the rate of change of the density to zero" is "useful in formalizing the notion that a system maintains its form over time" (Parr et al., 2019, p. 6). And the already quoted: "the internal and blanket states that constitute a subsystem are autopoietic, because their (nonequilibrium steady-state or ergodic) probability density is maintained by the flow of the subsystem's internal and active state" (Palacios et al., 2020, p. 5). The arrival at an unchanging set of relations describing the system in a NESS serves as the conceptual link between FEP and the properties its proponents associate with autopoiesis (self-organization, integrity, persistence, as we saw in Section 3).

Convergence to a NESS is an important point of contrast between FEP and enactive theory, one which few researchers have remarked upon. Like organizational approaches in biology (e.g., Longo & Montévil, 2014), enaction emphasizes the historical nature of life and mind (even more than the classical theory of autopoiesis does, see Di Paolo, 2018; Thompson, 2007; and Varela, 2011)[5]. Key technical enactive ideas, such as the concepts of adaptivity and sense-making, operate through mutually defined transformations of agent and environment that contradict the NESS condition. These ideas theorize precisely what happens in transitions between *different* relatively metastable situations (which might in some cases be approximately stationary in themselves) or in enduring situations where metastable conditions are not reached for significant periods. Varela uses the word "enaction" to refer precisely to such processes that happen "between one behavioral moment and the next" (Varela, 1991a, p. 106); and Di Paolo et al. (2017) describe such concrete acts of sense-making as the open transition between different microworlds of activity. In that sense, history is built into the very definition of the term *enaction*. Enactive ideas describe transitions between different regimes that typically require different dynamical descriptions, leading to changes in dynamical flows. In other words, these ideas concern what happens when a dynamical description suitable for one situation becomes unsuitable for the next; i.e., when we are no longer in a situation of *time translation symmetry*.[6] We can say that as a matter of definition, enaction focuses its theoretical apparatus on the *history* of mutual change between agents and their environments and for this reason enactive ideas are in tension with the *time-invariant* distribution of states that obtains in the NESS condition. We will argue next that this tension is in fact an incompatibility between the two approaches.

---

[5] In 1994, Maturana and Varela wrote individual prefaces to a new edition of *De Máquinas y Seres Vivos*, the first book on their joint work on autopoiesis. In his preface, Varela (2011) acknowledges that an important criticism of this early work was that autopoiesis seemed to imply a form of solipsism. Maturana and Varela had replaced the idea of mental representations with the idea that the environment triggers perturbations to the ongoing and operationally closed processes of the living system. This proposal included the notion of structural coupling, the idea that the living system and environment act as mutual sources of perturbation, triggering changes of state in each other. Varela notes that this notion seemed to many researchers to be a weaker alternative to the input-output approach of the information-processing metaphor, "because it seems to leave the phenomenon of interaction in a grey area of being a 'mere' perturbation" (ibid. 614). Structural coupling "does not properly take the account of the emerging regularities in the course of a history of interactions [...] Over these years I have developed an explicit alternative [...] turning the historical reciprocity into the clue of a co-definition between an autonomous system and its environment. I propose to call this point of view in both biology and cognitive science, *enaction*" (ibid.). History and organism-environment co-definition have been explicit concerns of the enactive approach since its beginnings, and they constitute a change in emphasis and orientation with respect to classical autopoietic theory.

[6] Time translation symmetry simply expresses that the *laws of change* describing a system remain unchanged if we apply a time displacement from t to $t + dt$. When talking about stationary regimes, distributions, densities, we use the term time-invariant, instead.

Let us first clarify what we mean by historicity in this context.[7] Historical processes and historical explanations abound in biology and the human sciences (e.g, Arthur, 1994; Byrge et al., 2014; Fischer & Bidell, 2006; Gould, 2002; Oyama et al., 2001; Thelen, 2005; Vygotsky, 2012; Waddington, 1957). We associate historical processes with qualitative changes in the space of possibilities for a given system. These changes go beyond changes in states. In dynamical systems terms, historical changes can be described as changes in parameters, whole sets of variables, in the laws of change, constraints, and boundary conditions that result in qualitative dynamical differences (e.g., a reshaping of the attractor landscape due to new (de)couplings to other systems; the transformation of variable sets due to changes in the body, the adaptation to novel external constraints that induce a reduction in degrees of freedom in the system, and so on). We also expect historical processes to be dissipative and often to inhabit extended critical regimes (with scale-invariant distribution of changes, i.e., not only manifested at longer timescales), and to be subject to changing nonholonomic constraints leading to phase spaces under persistent change. Phenomena that may be observed in historical processes can be hard to predict and include long-term memory, founder effects, locked-in retention, catastrophic forgetting, path-dependence, and broken symmetries of different kinds leading to diversity and variability (Desjardins, 2011, 2015; Di Paolo, 2001; Longo, 2018; Longo & Montévil, 2014; Montévil, 2020).

In sum and for our current purposes, historicity fundamentally entails broken time symmetries, i.e., changing laws, couplings, parameters, constraints, boundary conditions and variable sets. Systems in a stationary condition, in contrast, cannot accommodate this kind of historical change: they remain in their stationary condition unless something knocks them out of it and eventually forget their history due to their time invariant properties.[8] The unchanging density of states

---

[7] We follow authors such as Longo, Montévil, and others and use *historicity* to describe the historical character of some systems and processes, as elaborated in this paragraph. In other contexts, the term has a different meaning and refers mostly to the actuality or authenticity of historical events, persons, facts, etc.

[8] In this context, it is worth clarifying that models of active inference based on the FEP can be applied to the tracking and representation of nested sequential structures (Friston et al., 2018, 2020). These models can encode/generate relations in a hierarchy of nested sequences and in this way capture patterns over several timescales assuming a correlation between scale and model level. But the relation between these models and the question of historicity as defined above is, in our opinion, tenuous. These models describe the statistical relation of discrete nested sequences in order to generate a hierarchy of expectations to drive time-extended behavior, as in the case of reading a text, which may be described as organized hierarchically as sequences of sequences (letters, words, phrases, sentences, paragraphs, etc.). This does not entail that the system being modelled constitutes a historical process or that the model itself instantiates a historical process in the sense of changing spaces of possibilities and broken time-symmetries. We do not deny that deep temporal or sequential generative models may be useful in approximating historical changes, but they do so by tracking relations between nested sequences that are sufficiently close to being stationary. If the sequences change historically, say, as an infant transitions from crawling to walking (Adolph & Hoch, 2019), a nested sequential model tracking the muscle activation sequences of one kind of behavior does not by itself generate the new sequential relations needed

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

arrived at in FEP systems is definitionally time invariant. This is explicitly stated in the technical literature. For this reason, the relations between variable sets derived from such a condition cannot accommodate the possibility of the broken time symmetries that define historicity. For example, the appearance of a novel constraint induces novel correlations between variables and these new correlations provoke changes in the flow, thus making the system abandon its stationary condition. In other words, *the conservation of time invariant conditions cannot be guaranteed if time translation symmetry is broken.*

The tension between FEP and historicity has already been noted by Colombo & Wright (2021; see also Colombo & Palacios, 2021). Is this a minor tension or does it signal an incompatibility between the two approaches? We have already claimed that historicity can be said to be part of the very definition of the term *enaction*. In the rest of this subsection, we will show how important historicity is for enactive theorizing at all scales.

Historicity in life and mind can be found in many important cases that are central areas not only of enactive research but also of biology and psychology in general, such as the formation of sensorimotor habits (Di Paolo et al., 2017) and the formation of shared repertoires for coordinated action (Di Paolo et al., 2018). Historical processes are also manifested at shorter, behavioral and neural timescales, as in cases of soft assembly and critical agent-environment integration (e.g., Anderson et al., 2012), or the long-range coordination of neural populations at the moment of perceptual awareness (Varela et al., 2001). At multiple scales, living agents are constantly undergoing regulation in relation to the set of constraints that redefine their structural dynamics (phase spaces) moment to moment.

Evidence shows that in living and cognitive systems organizations can endure even if structures change. Without attempting to review the many cases in biological, neural, and cognitive systems where processes of historical transformation are at play, we can mention some examples. In biological processes: embryogenesis, life-cycle patterns, epigenetic variability, symbiosis, and metamorphosis.[9] At behavioral levels: fluid, critical agent-environment integration across scales in the development of perceptual learning, skill acquisition, expert tool use, and habit formation (Anderson et al., 2012). At the neurocognitive level: developmental plasticity, possibly many-many mappings between neural networks and cognitive functions (Pessoa, 2014), and more generally what Anderson (2014) calls "neural reuse," the continued acquisition of new uses for neural circuits in evolution and development without those circuits losing their original uses.

---

for the newer kind. At most, it can track them and reconstruct them a posteriori once the new situation is sufficiently stable.

[9] It is odd to find very few references in the FEP literature to these cases and the potential challenges they offer to the premises of the approach. One exception is Clark (2017), who realizes that cases such as metamorphic insects can be puzzling from a FEP perspective. The solution, he suggests, is to look at whole life cycles as the free energy minimizing strategy of complex organisms. This idea, however, challenges the assumption of convergence to a NESS, as Clark himself notes in a footnote (16).

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

Enaction attempts to accommodate many of these phenomena, so it is concerned with historical processes. Is historicity itself to be found at the core of its theoretical apparatus? The possibility of historicity is introduced unequivocally and explicitly in enactive theory through the concept of *adaptivity* (Di Paolo, 2005, p. 444). Adaptivity is the system's capacity to regulate its states and relation to the environment in ways that result in the avoidance of trajectories that move towards loss of viability. This definition of adaptivity entails that adaptive interventions necessarily modulate the dynamical landscape, for instance, by changing parametric relations to the environment, (un)coupling to other systems, or altering constraints or boundary conditions. These changes must necessarily result in an altered phase space of the adaptive system. By *definition*, this is the only way in which it is possible to change the tendencies of dynamical trajectories bound to break through the viability boundary in a state-dependent system. To elaborate our critical point, adaptivity, as defined and used in the enactive approach, implies broken time symmetries and so, as we have explained, is incompatible with assuming that a system remains in a NESS. Time invariant properties of the flow cannot be maintained if parameters, constraints, and even the system's variable sets are modified in the course of an adaptive event. Thus, two ideas that are key for each approach (adaptivity and NESS, respectively) contradict each other.

Adaptivity makes explicit the aspects of *time direction* and *time granularity* of agency (Di Paolo, 2005). An adaptive act is in itself *not* a conservation, but a modulatory deviation from an existing tendency that would lead to eventual loss of viability if left unchecked. An adaptive act takes time, has a particular time-course with different phases, and must occur within appropriate time constraints (given by requirements of speed, deadlines, by the relevant embodied and social norms, and so on). In important ways, things are just not the same before and after an adaptive intervention. History is made possible by these properties, especially as organisms become more complex, and adaptive acts relate to one another in networks of mutual influence and triggering, affecting, moment to moment, the relational constraints that shape the dynamics of the agent. The historicity of biological systems (Longo & Montévil, 2014) is precisely the ongoing broken symmetries in dynamical constraints, parameters, and even variable sets throughout a lifetime of changing mutual dependencies between organism and environment (including other organisms).

Historicity is not just present in enactive technical concepts but also elaborated in enactive research, particularly to theorize minimal agency, mastery of sensorimotor contingencies in action and perception, and the development of sensorimotor repertoires as the theoretical basis for the concept of sensorimotor agency and linguistic bodies (Di Paolo et al., 2017, 2018). One of the definitional requirements for agency proposed by enaction is that of *interactional asymmetry*. In dynamical systems terms, this requirement is meant to capture that not only are agent and environment coupled systems, but also that processes in the agent can trigger modulations (changes in parameters and constraints) of this coupling. Again, such modulations of parameters and constraints cannot guarantee that dynamical flows will remain in a NESS. They typically will not, and these changes are what allow the system to avert the potential loss of viability were it to remain in such a regime. As in the case of widespread extended critical transitions in biological systems (e.g., Longo & Montévil, 2014), the regulation performed in the acts of an agent can give rise to path-dependent changes on the constraints affecting the dynamical flows of the agent as a whole, essentially altering the configuration of the phase space of organism and environment, and even rendering notions such as attractive sets and NESS inadequate. The enactive concept of agency reaffirms for cognitive systems what Longo and Montévil predicate about biological systems: "Biological processes are more 'history based' than physical [non-biological] processes. Usual physical processes preserve invariants, whereas extended critical transitions [characteristic of biology] are a permanent reconstruction of organization and symmetries, i.e., of invariants" (ibid, p. 175).[10]

As far as human beings are concerned, evidence strongly suggests that processes of historical change cannot be disregarded. Ergodic and stationary conditions in psychology and neuroscience may sometimes make methodological sense in laboratory situations. But such conditions are not always made explicit and these assumptions can lead to systematic problems in interpreting the generality of empirical results or making group-to-individual inferences (Fisher et al., 2018). Indeed, several researchers interpret the current crisis of replication in psychology precisely as a crisis provoked by assuming ergodicity by default, a practice that be-

---

[10] It is important not to misread this quotation as stating that historical processes can occur only in biological systems. Broken symmetries occur in all kinds of physical systems, but usually in the form of single critical transitions. Longo and Montévil argue that biological systems are characterised by continuous extended critical transitions, and this is what makes them historical: "In our approach to biological processes as '*extended* critical transitions,' 'extended' means that *every point* of the evolution/development space is near a critical point. More technically, at the mathematical limit, the critical points form a dense subset of the multidimensional space of viability for the biological process. Thus, criticality is extended to the space of all pertinent parameters and observables (or phase space), within the limits of viability [...] In terms of symmetries, such a situation implies that biological objects (cells, multicellular organisms, species) are in a *continual transition between different symmetry groups*; that is, they are in transition between different phases, according to the language of condensed matter. These phases swiftly shift between different critical points and between different *physical determinations* through symmetry changes" (Longo & Montévil, 2014, p. 173). An example of extended criticality are the broken symmetries induced by the millions of mitosis events a multicellular organism undergoes every day where the resulting distributions of proteins, etc. are not identical and can lead to different integrated cellular dynamics which play a role in the development of different cell types and tissue differentiation. Dense sets of critical transitions make it difficult to explain biological systems via minimization or maximization principles: "As a further consequence of our approach, phylogenetic or ontogenetic trajectories cannot be defined by the geodesic principle. Indeed, they are not theoretically determined by invariants and their associated symmetries. Trajectories are continually changing in a relatively minor but extended way. Moreover, we expect the rate of these changes themselves not to be regular with respect to physical time, so that some temporal region can be 'calm' while others correspond [to a] sudden burst of changes" (ibid.). The existence of "critical surfaces" in relation to conservation of integration/viability can be shown in relatively simple models of information integration under environmental diversity (Aguilera & Di Paolo, 2019).

lies the lack of an underlying theory that would grant validity to the assumption for each given case Rose et al. (2013).

For instance, consider the formation of habits. We could say that there can be no habit without history, or that history is constitutive of the concept of habit. Enactive literature is very clear about this point: "To say that the habitual body acts as guarantee for the body at this moment is to say that one's lived body is a developmental being thick with its own history and sedimented ways of feeling, perceiving, acting, and imagining. These sedimented patterns are not limited to the space enclosed by the body's membrane; they span and interweave the lived body and its environment, thereby forming a unitary circuit of lived-body-environment" (Thompson, 2007, p. 33).

Importantly, history and contingency lead to diversity, which a theory of human bodies and minds cannot afford to leave without a proper theoretical grounding. The beginnings of such a theory are offered in (Di Paolo et al., 2018). Cultural and interpersonal variability in human beings are not statistical noise, but rather necessary consequences of what it means to undergo human becoming. Some events may enter the developmental, cognitive, and emotional history of a person as accidents, but if their effects endure or get amplified, this occurs by triggering a swerve in dynamical paths and reshaping dynamical landscapes. Fluctuations can become locked-in and change the course of subsequent history. None of this entails a loss of the system's integration, but it may entail transformations that redefine skills, sensitivities, meanings, and norms. Widespread biological and psychological phenomena of the kind FEP proponents intend to cover with their theory, (e.g., Veissière et al., 2019) are inherently historical.

The enactive approach provides accounts of perception and perceptual learning in terms of mastery of sensorimotor contingencies linking multiple scales from neural to developmental processes. Mastery, or acquisition of a know-how, is operationalized using a dynamical theory of equilibration that does not require systems to be in a NESS to work (Di Paolo et al., 2017). In fact, one of the phases of equilibration, accommodation, demands plastic changes in the agent and the environment such that the corresponding phase spaces are modified as a consequence, otherwise equilibration could not occur. Phenomena such as developmental spurts in skill level (Fischer & Bidell, 2006) are large scale manifestations of such changes in dynamical configurations (novel constraints, emergent parameters, changing variable sets). The variability entailed in changing dynamical configurations has been postulated as the origin of motor creativity (Orth et al., 2017), the very idea of which is rendered problematic without an account of historical change. All of these processes can be accommodated by the enactive theory of sensorimotor learning, which can also account for the existence of multiple developmental pathways, path-dependence, and the intrinsic variability found in all kinds of human skills (Adolph & Hoch, 2019; Kostrubiec et al., 2012; Thelen, 2005; Thelen et al., 1996).

A central claim of the enactive approach is that sense-making is the activity of an autonomous agent that contrasts with heteronomous information processing. To illustrate the sort of neural processes that underpin sense-making, Varela (1991a) and Thompson (2007) make repeated references to the work of Walter Freeman, who explicitly acknowledges history-dependence in the most basic neuroscientific scenario of stimulus processing: "The emerging [neural] pattern [...] is a state transition that is induced by a stimulus, followed by a construction of a pattern that is shaped by the synaptic modification among cortical neurons from prior learning. [...] Owing to dependence on history, the patterns created in each cortex are unique to each subject" (Freeman, 1999, pp. 149–150). Such history dependence of neural processes has been recognized as posing a methodological challenge of non-ergodicity for neuroscience (e.g., Medaglia et al., 2011).

Historicity is manifested over very different timescales, not just the longer timescales associated with development. In discussing the fine temporal structure of cognitive action, Varela proposes that coherent patterns of fast neural oscillations emerge at moments of concretion (significance) in action and perception (Varela, 1991a). This idea is empirically supported by studies of long-range neural (de)synchronization at moments of perceptual awareness (recognition) and action initiation, evidencing emergent processes that rapidly constrain and free dynamical flows contingently on the actions of the perceiver and her situation (Varela et al., 2001). This perspective on brain function puts the emphasis on neural processes radically altering their dynamics in fluid and adaptive ways so as to meet the demands of a concrete sociomaterial situation (e.g., Anderson, 2014; Fuchs, 2017; Pessoa, 2014). Such rapid moments of high dimensional dynamic expansion followed by lower dimensional coordination is one of the meanings given to the term *enaction* itself, as we have said. Accordingly, "the hinge that articulates enaction consists of fast non-cognitive dynamics wherein a number of alternative microworlds are activated. These hinges are the source of both common sense and creativity in cognition" (Varela, 1991a, p. 109). The latter remark indicates that these fast neural processes are not just history-*dependent* ("common sense") but also history-*making* ("creativity").

The idea of inhabiting and transitioning between microworlds has been further developed by Di Paolo et al. (2017), where the authors discuss the possibility of extended periods of high-dimensional exploration of sensorimotor possibilities when the transition between activities does not occur or for whatever reason is arrested. Examples include periods of hesitation or confusion as to the relevant activity one is engaged in, either by lack of determination (not knowing exactly what comes next out of several options) or overdetermination (current activity continues pre-reflectively while we are reflectively aware that it is no longer the relevant activity). Extended exploratory periods that do not settle into metastable modes of activity also include situations in which the required adaptations to some radical sensorimotor disruption have not yet been achieved (changing habits, experiments involving sensorimotor alterations, major injury, trauma, etc.). The authors call these moments of uncommitted dynamical engagement the zero-mode of activity. Using a network metaphor to describe the functional and structural relations found within a repertoire of sensorimotor schemes, the authors remark "through regional developmental differentiation and integration of schemes, integrated subnetworks can be formed that correspond well to notions such as activities, microworlds, and sensorimotor genres. The structure of the sensorimotor network reflects the *history of the agent*. In humans the possibilities are open-ended and path-dependent, as we would expect, leading to a way of characterizing otherwise vague concepts such as sensorimotor styles" (Di Paolo et al., 2017, pp. 177–178).

In sum, whether at biochemical, neural, behavioral, or developmental scales, enactive ideas all share in common the expectation that biological and cognitive systems are historical, changing in path-dependent ways, not only at some points in time, but regularly, at different scales, and throughout their lifetime. For this reason, historicity is definitionally built into its key concepts (enaction, adaptivity, agency, sense-making). History is also a thread that drives current directions of the enactive approach (e.g., see recent articulations of the notion of open-ended human becoming in De Jaegher, 2021; and Di Paolo, 2021).

Could this incompatibility between key ideas in each approach be turned into a reconcilable tension, i.e. a challenge for FEP? A possible attempt might take the form of an argument that renders "historical" changes as only apparent (i.e., not really involving broken time symmetries) and explains them as transitions between existing and pre-established "micro-regimes" within what is otherwise a system in a NESS (when seen at appropriately long scales). This is similar to Clark's (2017) idea to regard whole life cycles as a free energy minimizing strategy. Thus a transition between one microworld and another could be described as a switch between metastable "micro-regimes" within the larger, lifetime NESS. This explanation could accommodate some cases of switches between routine activities, but if it were generally the case, it would imply that all of our activities are routine, statistically speaking. The stationary condition must translate into a steady distribution of transits between "micro-regimes." Engaging in novel activities (e.g., learning to drive) or abandoning old ones (because of loss of interest or other factors), however, cannot be explained in this way. The idea of a whole life cycle characterised by a NESS does not seem applicable to the open-endedness of the human case.

Another way a FEP account might accommodate historicity could be by not assuming that the system overall must behave in a stationary manner, but that it frequently inhabits different steady-state "micro-regimes" and sometimes is able to transit from one to another without the whole set of steady states being predetermined or fixed over time. This would free the account from the requirement that transitions between "micro-regimes" must also follow a stationary distribution because stationarity is only required locally within a "micro-regime." But such a refinement still would not answer the question of what exactly happens during such transitions, what triggers them, how frequent they are, how long they are expected to last, what norms regulate the adaptive mechanisms at play in them (since free energy minimization is not guaranteed during a transition), and so on.

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

Moreover, while this idea may bring the FEP closer to some aspects of history, it would not be suitable for the extended criticality and dense transitions sets postulated by Longo & Montévil (2014) as characteristic of biological systems, which in some cases exhibit symmetry breaking of the kind that would not even guarantee steady-state "micro-regimes."

We repeat our point that given the current state of affairs, and the respective importance of the NESS condition for FEP and of historicity for enaction, this tension counts as a theoretical incompatibility (in the sense that one cannot assert both simultaneously). If new variants of FEP were developed replacing the need for the system to converge to and remain in a NESS, the situation might change. But the problem of accounting for the historicity of life and mind cannot be solved by admitting that the need for a NESS condition is merely an idealization that can sometimes be relaxed or is relative to a timeframe of relevant observation (e.g., Wiese & Friston, 2021, p. 7). To admit this is, again, to concede that the FEP does not have the claimed status of a universal principle, since its domain of applicability would remain limited to relatively simple cases. Such a move begs the question of what happens to biological and cognitive systems at the moments the steady-state assumption does *not* hold, moments that, as we have argued, are copiously found at all scales of biological and psychological phenomena. A theoretical biology and a theoretical psychology should be able to answer when this assumption is *sufficiently valid* for its application to be useful *and* what happens when it is not met.

If important phenomena (e.g., ontogeny, acquired immunity, the time course of illness or injury and recovery/compensation, skill acquisition, habits, development towards emotional maturity, trauma, personality, language, abstract thought, social interactions, expert use of tools, cultural history, evolution of technology, changes in modes of production, and innumerably other path-dependent processes) fundamentally break the time invariance of NESS, it does not follow that we should discard FEP ideas as useful research tools, provided we can specify the conditions under which their use makes sense. But, as currently formulated, FEP does not answer this question by itself. It would be as if we were expected to deduce the shape of the Earth from a theory that assumes its surface must be flat just because under some conditions this is a good approximation but without specifying what these conditions are or what happens when they are not met. The necessary specifications must come from a theoretically-loaded account that tells us when to expect assumptions that lead to a NESS to hold and when not to expect them. Elaborations to the existing formulation of the FEP may provide some answers, but they may require seeking different, perhaps less restrictive starting points. It goes without saying that the same theory cannot adopt an assumption as fundamental and deduce when it does not hold as a corollary.

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

4.2 Mutual enablement and co-constitution of internal and external domains

In Section 3, we noted the recurrent and puzzling confusion between Markov blankets and what autopoietic and enactive theories specify as operationally closure, and in particular processes of organizational self-distinction. These processes of self-distinction do occasionally—but not exclusively—take the form of processes of structural boundary formation, regulation, and maintenance. They can also take other forms (e.g, immune responses against invading pathogens, temperature regulation in warm-blooded animals, musculoskeletal systems of tension and compression maintaining bodily shape). In all instances they involve adaptive regulations of exchanges between organisms and the environment. We can ask whether the confusion between such processes and Markov blankets is symptomatic of another tension between the two approaches, a tension in how they conceive of relations between organism and environment and relations between different organisms; whether these relations are merely informational (i.e., contextual), or can also be transactional (e.g., mutually enabling) or even constitutive of biological and cognitive/affective processes. Enactivists assert a strong notion of world-involvement, i.e., processes in the environment play more than informational roles in the constitution and actualization of life and mind (Di Paolo et al., 2017; Thompson & Stapleton, 2009). To enact a world of significance is to engage in actual acts, which are material events with spreading consequences that are both world-changing and agent-changing. Environmental and biological/cognitive processes are mutually enabled and mutually constituted. They interpenetrate at all scales and they coordinate across scales. Can these claims be accommodated by the idea of Markov blankets?[11]

Here we will re-examine the concept and processes of self-distinction and why these processes should not be quickly equated with Markov blankets. We will argue that many important exchanges between organisms and the environment do not follow the statistical constraints of Markov blankets and can in fact be unmediated or unconditioned, only to be dealt with adaptively after the fact. After this, we turn to what enactivists describe as world-involvement, i.e., the ways in which relational processes involving both organism and environment play enabling and constitutive roles in the activity of an agent as well as in its ongoing individuation. These relational processes affect the operation of this activity *and* contribute to establishing the norms of viability that this activity follows. In particular, they help establish which encounters can or should be mediated by states of the organ-

---

[11] For recent critical discussions about Markov blankets, see Bruineberg et al. (2021) and Raja et al. (2021). Bruineberg and colleagues argue that a simple statistical idea applied in Bayesian networks and describing how variables may be shielded from variations in other variables has been made to do some heavy conceptual lifting in the FEP framework, being used, as we discuss in this paper, to play the role of boundaries and sensorimotor interfaces. Raja and colleagues, similarly note this problem and the fact that the drawing of a Markov blanket is often ad hoc, rather than following a consistent method.

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

ism. For this reason, world-involving relational dynamics operate at a level prior to Markov blankets and cannot be established by them. In other words, these two points indicate that Markov blankets insufficiently specify what enactivists highlight about the relation between organism and environment. It is possible, however, that the tensions we discuss in this subsection may take more the form of a challenge than that of a strict incompatibility, as we shall see.

Historicity and the co-constitution of organism and environment are internally related in the enactive approach. Concerns about the conservation of organization are mostly linked to the self-production requirement of autopoiesis (the regeneration of the conditions that continuously give rise to the operationally closed network of processes making up the organism). Concerns about barriers, boundaries, and in general about an organism's relation to its environment are mostly linked to the condition of self-distinction in the definition of autopoiesis. From an enactive perspective, self-distinction and self-production are dialectically related (Di Paolo, 2018), that is, they are mutually dependent, though distinct, moments of autopoiesis (Thompson, 2007; Varela, 1991b). You cannot have one set of processes and not the other as long as the organism lives, yet the processes are not the same. All processes subserving self-distinction are themselves products of self-production. In contrast, Markov blankets in FEP systems are there by assumption.[12]

In other words, Markov Blankets are not *produced* by the system the same way that self-distinction processes and structures are. Hence there is nothing in the Markov Blanket that necessarily links it to processes of organismic constitution. This is a discrepancy also noted by Raja et al. (2021), who argue that the choice of where a Markov blanket should be is rather *ad hoc* and follows the convenience of each case (see also Bruineberg et al., 2021). This is not a problem in itself; it might even be an advantage in some cases. But it points to a difference between FEP and enaction.

---

[12] The assumption is considered self-evident: "Clearly, one needs to differentiate between the system and its environment [...] To do this, we have to introduce a third set of states that separates internal from external states. This is known as a Markov blanket." (Ramstead et al., 2018, pp. 3–4). This explains the wide applicability of Markov Blankets in examples and models of FEP, sometimes taking the role of boundaries, other times of action and sensory states, and so on. As Raja et al. (2021) argue the choice of where a blanket is located seems to be relatively arbitrary and made to fit the convenience of an FEP interpretation. In an attempt to provide a more principled interpretation of Markov Blankets, Friston (2019) associates them with the structured (ordered) flows in far-from-equilibrium dissipative structures (Nicolis & Prigogine, 1977). But he acknowledges that this poses a problem for the FEP formalism, because such structures are in constant material and energetic flux, which contradicts the assumption of a random dynamical attracting set, and confirms that the two central assumptions of FEP are independent, unlike self-production and self-distinction for enaction. Friston considers this problem an unresolved challenge for FEP and its generalization to so-called wandering sets (which would not suffice to address the cases of historicity discussed in this section, as these include the possibility of changing sets of variables, such as emerging novel agent-environment relations). The whole formalism, as it currently stands, rests on "the simplifying assumption that over a suitable time scale, blanket states are well defined—as a subset of attracting states" (Friston, 2019, p. 50).

We may still ask whether Markov blankets, even if they are not produced by the organism itself, could still provide a sufficiently informative description of the organism-environment relation. In response to this question, we find that self-distinction is not sufficiently specified by a statement about statistical dependencies or sparsity of connections between variables, nor does it entail any such statement. Let us see why.

We confront a semantic difficulty at this point. A Markov blanket helps define the sets of internal and external variables in the FEP. What counts as internal and external variables in the enactive approach is less straightforward. It is in principle possible to identify what processes belong or do not belong to the operationally closed network of the organism (though in practice there may be some ambiguities; consider for instance the layer of warm air that surrounds the skin or fur of warm-blooded animals). But this distinction does not translate into a straightforward topological relation as might be postulated by pointing to boundary processes such as membranes, cell walls, skin, etc. The network of operationally closed processes can extend beyond these boundaries and environmental processes can transverse them. There is a semantic difficulty in comparing apparently simple terms such as "internal" and "external" in the two theories. From an enactive perspective and for the purposes of this discussion, we propose to treat as internal those variables that are topologically contained by an organism's boundary surfaces in cases where these are well-defined by boundary processes of self-distinction, and as external those variables that lie outside boundary surfaces. The meaning approaches the usage in the FEP literature but is not exactly the same, because this distinction between internal and external does not exactly map to that between organism and environment in the enactive view, nor does it presuppose any statistical relation between internal and external variables.

Keeping this in mind, we repeat that self-distinction is not supported structurally only by the formation of organismic boundaries. "Non-boundary" processes of self-distinction include the immune system, CRISPR-Cas anti-virus defense, sweating, shivering and other processes that sustain temperature regulation, processes of maintenance of shape, such as cytoskeletons, musculoskeletal tensegrity, tissue formation and regulation of the extracellular matrix, and so on. In other words, they include internal processes (in the sense just defined). Because these "non-boundary" self-distinction processes do not lie at the topological interface between organism and environment, they do not generally entail any particular kind of conditioned independence between internal and external variables; on the contrary, they are regulations of encounters between internal and external processes, where internal effects can follow external factors quite directly. Even if they operate internally, these processes of self-distinction adaptively regulate the relation between agent and environment. However, because they operate internally, they seem difficult to describe by Markov blankets.

Let us consider "boundary" processes of self-distinction, such as cell membranes. We find a large number of processes that cross boundaries and barriers in both directions without any statistical mediation. For example, osmosis and other forms of diffusion, the ergodicity-breaking formation of protein nanoclusters affecting ion-channels in neurons (Weron et al., 2017), horizontal transfer of plasmids in bacteria (Soucy et al., 2015), the cumulative effects of diet and adaptation to toxins or their expulsion (Landecker, 2011), direct energy transfer in photosynthesis, temperature equilibration and temperature-induced neural plasticity in cold-blooded animals (e.g., Beltrán et al., 2021), and "externalized" physiological processes such as the extraction of gaseous oxygen underwater by insects trapping air bubbles in their abdomen (e.g., Turner, 2000). Not to mention more complex processes such as cell division, sexual reproduction, host-parasite relations, gut microbiota, and symbiosis in general. The list does not include external processes that are hard or impossible to regulate, such as the effects of ionizing radiation, gravity, or unavoidable physical impacts, but that nevertheless affect internal variables directly. External processes affect internal variables without mediation in all of these cases.

This range of exchanges between internal and external processes does not pose a problem for enaction. Some of these exchanges are avoided or rejected, some are mediated by the activity of the organism, others remain unmediated in the encounter and regulated only in their effects. The condition of operational closure is not necessarily lost even in the latter case, provided the effects of an unmediated exchange can be adaptively assimilated (e.g., a fever response to bacterial infection). These exchanges, however, do not require that internal and external variables relate through a conditional statistical independence and in fact in many of the cases mentioned, this will not be the case.

We suggest that processes of self-distinction, as conceived in the enactive approach, do not generally behave like Markov blankets do. Agency, in the enactive view, involves not just the regulation of boundaries, but also the *a posteriori* regulation of more direct, boundary-crossing influences of the world, by taking advantage of these influences when useful (e.g.. nourishment) and compensating for their negative effects when not (e.g., fever response). From an enactive perspective, Markov blanket conditions are too strong and too undiscriminating to account for the complexity of the organism-environment relation, and this is a point of tension.

A related point of tension concerns *world involvement*, a term enactivists sometimes use to capture the relational aspects of the agent-environment coupling (Di Paolo, 2014; Di Paolo et al., 2017). World involvement stresses the fact that the environment is involved in what an agent is and what it does in more than contextual ways, that is, its involvement goes beyond being a source of information. The world is also a material enabler of cognitive acts and may also play constitutive roles in such acts, particularly in contributing to determine their normative aspects (what counts as an act succeeding or failing, what makes an act preferable to another).[13]

---

[13] The distinction between contextual, enabling, and constitutive factors was introduced by De Jaegher et al. (2010) to ask the question of whether the dynamic patterns that we observe in

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

The world enables and constitutes embodied agency over and above any information it provides to the cognitive processes of the agent. This is not to say such information is irrelevant (provided we describe clearly what we mean by this term, and provided we can establish what counts as the frame of relevance in each case, something that can change historically). It means that other, non-informational, aspects of material processes are also enabling and constitutive factors in cognitive acts and in sustaining the autonomy of the agent. This is what world involvement means. Going on a diet, training for a new skill, and so on—the environmental processes in such activities do not merely inform our bodies through sensory-screened inputs about how they should change as a consequence, more often than not they *change* our bodies directly. Drinking a glass of water quenches our thirst in a way that is different from encountering a useful piece of information. A zero-G environment does not inform the body of a cosmonaut that it should grow less dense, it "makes it" do so.

In addition to these enabling and constitutive roles in the activity and ongoing self-constitution of the agent, the relational processes of world involvement also contribute to the establishment of the norms of an agent's activity, as well as its conditions of viability. For instance, by engaging in a novel manual skill, we learn through confronting obstacles to accommodate new norms that emerge from these material encounters, as in the different ways an expert potter prepares different kinds of clay. We learn to establish a new frame of relevance as a historical transformation of our bodies and our relation to the environment. This frame is necessary to determine what counts as meaningful information (e.g., cues we used to be indifferent to, but now tell us whether this is good wood for sawing, or well-prepared clay for shaping a pot, or the age of a wine from its bouquet). Relevance, meaning, norms are irreducibly relational properties. Because they establish what counts as informational for the agent, these material and relational transformations are prior to any informational account we can provide.

Both of these points—the enabling and constitutive roles environmental processes play in sense-making and in agent constitution, and the role they play in establishing norms for the activity of an agent—indicate tensions with the idea of Markov Blankets. These tensions, however, partly rely on how we choose to define internal and external processes. For this reason, it might be possible to work on interpretations of enaction and FEP that avoid these apparent contradictions, so these tensions are at least *challenges* to the FEP; we do not claim they are incompatibilities.

---

social interactions can (sometimes) be said to be constitutive of social cognitive performance, i.e., whether there is social intelligence in the dynamics of the interaction itself and not just within the interactors. A contextual factor merely modulates a phenomenon under observation; it changes it, but is not required for the phenomenon to exist or to be the kind of phenomenon that it is. An enabling factor is one without which the phenomenon would not occur. A constitutive factor is one without which the phenomenon would not be the kind of phenomenon that it is (see also De Jaegher et al., 2016).

These are not minor challenges, though. They can lead to divergences in how each approach conceives action and perception. The enactive approach postulates a theory of sensorimotor mastery (know-how) based on equilibrated relations between sensorimotor schemes (Di Paolo et al., 2017). These schemes are normatively integrated flows of coordinated body and environmental processes in coupling (e.g., the movement of water from a glass into the mouth, the activation of swallowing muscles, its flow within the body, and so on). They combine internal and external processes. Because traffic across a Markov Blanket is conceived as mediated by sensory and effector states there is a tendency of conceiving of the organism-environment relation in informational terms and this is verified by terms such as evidence, surprise, inference, etc. This promotes (not necessarily compels) internalist views of sensorimotor know-how that leave out precisely the transformative effects of the world on our bodies. An informational interpretation assumes that such effects have already occurred, such that the norms that determine what counts as relevant information are already in place and the relevant distinctions already established. But explaining how we can reach this stage is what enactive theories of action and perception are about.

For a theory of perception inspired by free energy minimization, such as Seth's (2014), embodied factors such as sensorimotor contingencies play a role in the informational economy of hierarchical predictive models, structuring data, selecting sampling strategies, and generating error-based corrections. Such proposals remain all "in-the-head" and reduce the rich materiality of the worldly constituents of perceptual experience to error signal generation (Di Paolo, 2014). For the enactive approach, in contrast, it is from the material constraints introduced by world-involvement in action and perception that the norms of such activities arise (what to do? what counts as doing it right?). These emerging norms result from agent-environment transactions and crucially frame what counts or does not count as relevant information for the agent and not for the external observer; what is relevant and what is not, what needs to be optimized and what may be safely ignored. In this sense, the mastery of these norms, a transactional process not bound by the brain, is constitutive of perceptual experience (Di Paolo et al., 2017). Whatever one may think of the merits of these theories, the fact remains that they are different in important aspects (this is sometimes acknowledged by those who would still try to reconcile them; see Clavel Vázquez, 2020).

5 Conclusions

It is important that our motivation in this article should not be misconstrued as a general statement against intellectual synthesis or against the cross-fertilization of ideas. Enactivists generally welcome and celebrate such exchanges. Nor is our discussion an exercise in gatekeeping. The enactive approach has undergone important changes over the last decades and will continue to evolve. In part, these changes have resulted from the meeting (sometimes the confrontation) of enactive ideas with other theories, disciplines, and practices. A case in point is the ongoing discussion of the relations between enaction and ecological psychology (Chemero, 2009; Di Paolo et al., 2021).

Nevertheless, the process of inter-theoretical comparison cannot be bypassed or replaced by rhetorical devices, such as stating that some similarities are “fairly straightforward” or by cherry-picking what to keep and what to ignore about a theoretical framework.

Attempts at discussing the compatibility or complementarity between different theories are initially motivated by what look like interesting similarities and resonances between them. This is fine. But the work of inter-theoretical comparison cannot stay at the level of pointing out these similarities. It requires critical examination to determine whether similarities are only apparent and whether they are not based on misinterpretations. This work demands a deeper and detailed engagement with the assumptions behind the theories being compared, as well as their contexts and goals. Despite its abundance, the literature defending compatibilities between the FEP and autopoietic and enactive theories has advanced very little beyond bald statements of apparent similarities between these approaches. Such statements remain at the beginning of the work that needs to be done.

In this article, we have shown that once we take the required steps of checking for misinterpretations and comparing fundamental assumptions, the FEP, as it stands formulated today, presents a series of tensions with the theories of autopoiesis and enaction. Not only do we find misreadings—such as confusing statements about organization with statements about structures, or conflating notions of self-distinction with material boundaries and the latter with Markov blankets—but we also suggest that at the root of these misreadings may lie conflicting views about the constitution of agency and self-individuation in biological and cognitive systems.

Our conclusion is that, for these reasons, the FEP—at least in its present form—and the enactive approach are not as compatible as has been suggested. The inter-theoretical differences we have shown are not minor, and spread respectively through each project as a whole. In some cases, we argued that apparent contradictions can be shown to lead to incompatible statements and claims. In other cases they at least indicate a challenge for proponents of FEP interested in connecting to enactive theory to develop extensions, clarifications, or make explicit how the tension can be resolved.

In particular, we have shown that enaction conceives of bodies and sense-making as irreducibly historical (both in the sense of depending on the past and in the sense of changing, or enacting, the present) and co-constituted by the sociomaterial world in ways that go far beyond the processing of information. The centrality of the historical transformation of agents and worlds for the enactive approach is in plain view and a definitional matter in its technical concepts. It is also a direct reading of its classical slogan: *Laying down a path in walking* (Thompson, 2007; Varela et al., 1991). In contrast, the key assumptions of the FEP, such as

NESS, make it difficult for it to reach a similar conception of life and mind, as we have shown. Because of this, we consider this to be an incompatibility between the two approaches. Other tensions turn around how each theory conceives of the relation between agent and environment. We have suggested that Markov blankets do not seem to describe the possible interactions that can be accommodated by the enactive idea of operational closure, particularly those that take place "internally." However, this tension revolves around tricky semantic issues such as definitions of seemingly obvious terms such as "internal" and "external." If semantic interpretations were modified, then this tension might turn out to be simply a challenge and not an incompatibility, so for now, we take it as such.

There are other differences between the two approaches that we have not explored here. For example, there are aspects of the temporality of sense-making and lived experience apart from historicity that can be said to be in tension with notions such as predictions and active inference. Sense-making by its technical definition is future-oriented activity, but this orientation can take a variety of forms, from general states of readiness, trust, or alert and their opposites, to conditions of commitment to action and focus or states of openness at the transition between activities. Of course, sometimes sense-making can also take the form of expectations, inferences, and predictions. But the varieties of future orientation do not reduce to the latter cases only. Establishing this point, however, would require careful conceptual, phenomenological, and terminological analysis, which we cannot pursue here.

Another difference we have not explored concerns the social dimension. Human bodies and minds, apart from being constitutively historical and diverse, are irreducibly social. Any approach that hopes to provide explanations for human cognition must acknowledge this fact and supply the right theoretical articulations to work with it. Enactivists have proposed the concept of participatory sense-making (De Jaegher & Di Paolo, 2007) as the keystone for developing an enactive approach to social, linguistic, and ethical agency (Di Paolo & De Jaegher, 2021). This idea, in our view, is at odds with the notions of blanketed individual persistence of cognitive individuals at the heart of the FEP. In participatory sense-making, actions and intentions become socially constituted and transformative of individual agency; they involve not only other participants but also the relational dynamics of the interaction (De Jaegher et al., 2010). We suggest that it would be difficult to account for these mutually transformative processes if each interactor were subject to the conditional statistical independence that is forced on them by Markov blankets and the need to remain in a NESS during the participation process. Again, a full argument, which space prevents us from providing here, would be necessary to establish this point and decide whether it is only a challenge or an incompatibility.

We have not dwelled on the question of what motivates researchers to attempt to make the FEP compatible with enaction and autopoiesis. One driving motivation seems to be the need to provide more embodied and less neurocentric interpretations of free energy minimization and active inference. Although this is a good motivation, we think that elaborating this interpretation is an uphill task, in no small measure because of the neurocentric and computational origins of FEP and predictive processing. A lot of intellectual effort in the literature that we have discussed is spent, almost apologetically, in revisionist attempts at reinterpreting computationally loaded ideas, such as model, inference, and prediction, in terms of embodiment, dynamics, and agent-environment couplings. Having a theoretical approach at hand, such as enaction, that already starts from an embodied and situated perspective could perhaps have given the impression that the uphill task was achievable by establishing a few bridges between the theories. We have argued that things are not so easy. The styles of thinking are different and the ideas difficult to harmonize. Whether there are other ways to make the FEP coherently more embodied is not a question we have set ourselves to answer here.

Acknowledgments
We are grateful to Miguel Aguilera, Maxwell Ramstead, and the reviewers of this paper for their comments and suggestions. We also thank Wanja Wiese for his editorial advice and support during the review process.

References

Adolph, K. E., & Hoch, J. E. (2019). Motor development: Embodied, embedded, enculturated, and enabling. *Annual Review of Psychology*, *70*(1), 141–164. https://doi.org/10.1146/annurev-psych-010418-102836

Aguilera, M., & Di Paolo, E. A. (2019). Integrated information in the thermodynamic limit. *Neural Networks*, *114*, 136–146. https://doi.org/10.1016/j.neunet.2019.03.001

Aguilera, M., Millidge, B., Tschantz, A., & Buckley, C. L. (2021). How particular is the physics of the free energy principle? *Physics of Life Reviews*. https://doi.org/10.1016/j.plrev.2021.11.001

Allen, M., & Friston, K. J. (2018). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese*, *195*(6), 2459–2482. https://doi.org/10.1007/s11229-016-1288-5

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain.* MIT Press.

Anderson, M. L., Richardson, M. J., & Chemero, A. (2012). Eroding the boundaries of cognition: Implications of Embodiment. *Topics in Cognitive Science*, *4*(4), 717–730. https://doi.org/10.1111/j.1756-8765.2012.01211.x

Arthur, W. B. (1994). *Increasing returns and path dependence in the economy.* The University of Michigan Press.

Ashby, W. R. (1962). Principles of the self-organizing system. In H. V. Foerster & Jr. G. W. Zopf (Eds.), *Principles of self-organization: Transactions of the University of Illinois Symposium* (pp. 255–278). London: Pergamon Press.

Baltieri, M., Buckley, C. L., & Bruineberg, J. (2020). Predictions in the eye of the beholder: An active inference account of watt governors. In J. Bongard, J. L. L. Hebert-Dufrsne, R. Dasari, & L. Soros (Eds.), *Proceedings of ALIFE 2020: The 2020 conference on artificial life* (pp. 121–129). MIT Press. https://doi.org/10.1162/isal_a_00288

Barandiaran, X. E., Di Paolo, E. A., & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, *17*(5), 367–386. https://doi.org/10.1177/1059712309343819

Beer, R. D. (2015). Characterizing autopoiesis in the game of life. *Artificial Life*, *21*(1), 1–19. https://doi.org/10.1162/artl_a_00143

Beer, R. D. (2020). Bittorio revisited: Structural coupling in the game of life. *Adaptive Behavior*, *28*(4), 197–212. https://doi.org/10.1177/1059712319859907

Beltrán, I., Herculano-Houzel, S., Sinervo, B., & Whiting, M. J. (2021). Are ectotherm brains vulnerable to global warming? *Trends in Ecology & Evolution*, *36*(8), 691–699. https://doi.org/10.1016/j.tree.2021.04.009

Biehl, M., Pollock, F. A., & Kanai, R. (2021). A technical critique of some parts of the free energy principle. *Entropy*, *23*(3), 293. https://doi.org/10.3390/e23030293

Bitbol, M., & Luisi, P. L. (2004). Autopoiesis with or without cognition: Defining life at its edge. *Journal of The Royal Society Interface*, *1*(1), 99–107. https://doi.org/10.1098/rsif.2004.0012

Bourgine, P., & Stewart, J. (2004). Autopoiesis and cognition. *Artificial Life*, *10*(3), 327–345. https://doi.org/10.1162/1064546041255557

Bruineberg, J., Dolega, K., Dewhurst, J., & Baltieri, M. (2021). The emperor's new Markov blankets. *Behavioral and Brain Sciences*, 1–63. https://doi.org/10.1017/s0140525x21002351

Bruineberg, J., & Hesp, C. (2018). Beyond blanket terms: Challenges for the explanatory value of variational (neuro-) ethology. *Physics of Life Reviews*, *24*, 37–39. https://doi.org/10.1016/j.plrev.2017.11.015

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2018). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese*, *195*(6), 2417–2444. https://doi.org/10.1007/s11229-016-1239-1

Byrge, L., Sporns, O., & Smith, L. B. (2014). Developmental process emerges from extended brain–body–behavior networks. *Trends in Cognitive Sciences*, *18*(8), 395–403. https://doi.org/10.1016/j.tics.2014.04.010

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

Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy*, *53*, 3–27. https://doi.org/10.1111/sjp.12120

Clark, A. (2017). How to knit your own Markov blanket: Resisting the second law with metamorphic minds. In T. Metzinger. & W. Wiese (Eds.), *Philosophy and predictive processing* (pp. 1–19). Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958573031

Clavel Vázquez, M. J. (2020). A match made in heaven: Predictive approaches to (an unorthodox) sensorimotor enactivism. *Phenomenology and the Cognitive Sciences*, *19*(4), 653–684. https://doi.org/10.1007/s11097-019-09647-0

Colombo, M., & Palacios, P. (2021). Non-equilibrium thermodynamics and the free energy principle in biology. *Biology & Philosophy*, *36*(5). https://doi.org/10.1007/s10539-021-09818-x

Colombo, M., & Wright, C. (2021). First principles in the life sciences: The free-energy principle, organicism, and mechanism. *Synthese*, *198*(S14), 3463–3488. https://doi.org/10.1007/s11229-018-01932-w

Constant, A., Clark, A., & Friston, K. J. (2021). Representation wars: Enacting an armistice through active inference. *Frontiers in Psychology*, *11*, 598733. https://doi.org/10.3389/fpsyg.2020.598733

Costa, L. D., Friston, K., Heins, C., & Pavliotis, G. A. (2021). Bayesian mechanics for stationary processes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *477*(2256), 20210518. https://doi.org/10.1098/rspa.2021.0518

De Jaegher, H. (2021). Loving and knowing: Reflections for an engaged epistemology. *Phenomenology and the Cognitive Sciences*, *20*(5), 847–870. https://doi.org/10.1007/s11097-019-09634-5

De Jaegher, H., & Di Paolo, E. A. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507. https://doi.org/10.1007/s11097-007-9076-9

De Jaegher, H., Di Paolo, E. A., & Adolphs, R. (2016). What does the interactive brain hypothesis mean for social neuroscience? A dialogue. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150379. https://doi.org/10.1098/rstb.2015.0379

De Jaegher, H., Di Paolo, E. A., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. https://doi.org/10.1016/j.tics.2010.06.009

Desjardins, E. (2011). Historicity and experimental evolution. *Biology & Philosophy*, *26*(3), 339–364. https://doi.org/10.1007/s10539-011-9256-4

Desjardins, E. (2015). Historicity and ecological restoration. *Biology & Philosophy*, *30*(1), 77–98. https://doi.org/10.1007/s10539-014-9467-6

Di Paolo, E. A. (2001). Artificial life and historical processes. In J. Kelemen & P. Sosík (Eds.), *Advances in artificial life* (pp. 649–658). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44811-x_75

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*(4), 429–452. https://doi.org/10.1007/s11097-005-9002-y

Di Paolo, E. A. (2009). Extended life. *Topoi*, *28*(1), 9–21. https://doi.org/10.1007/s11245-008-9042-3

Di Paolo, E. A. (2014). The worldly constituents of perceptual presence. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00450

Di Paolo, E. A. (2018). The enactive conception of life. In A. Newen, S. Gallagher, & L. de Bruin (Eds.), *The Oxford handbook of cognition: Embodied, embedded, enactive and extended* (pp. 71–94). Oxford: Oxford University Press.

Di Paolo, E. A. (2021). Enactive becoming. *Phenomenology and the Cognitive Sciences*, *20*(5), 783–809. https://doi.org/10.1007/s11097-019-09654-1

Di Paolo, E. A., Buhrmann, T., & Barandiaran, X. (2017). *Sensorimotor life: An enactive proposal.* Oxford University Press.

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic bodies: The continuity between life and language.* MIT Press.

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187

Di Paolo, E. A., & De Jaegher, H. (2021). Enactive ethics: Difference becoming participation. *Topoi*. https://doi.org/10.1007/s11245-021-09766-x

Di Paolo, E. A., Heras-Escribano, M., Chemero, A., & McGann, M. (Eds.). (2021). *Enaction and ecological psychology: Convergences and complementarities*. Frontiers Media SA. https://doi.org/10.3389/978-2-88966-431-3

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

Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action, thought, and emotion. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology, vol. 1: Theoretical models of human development* (pp. 313–399). New York: Wiley.

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Freeman, W. J. (1999). Consciousness, intentionality and causality. *Journal of Consciousness Studies*, *6*(11–12), 143–172. https://www.ingentaconnect.com/content/imp/jcs/1999/00000006/f0020011/995

Friston, K. (2013). Life as we know it. *Journal of The Royal Society Interface*, *10*(86), 20130475. https://doi.org/10.1098/rsif.2013.0475

Friston, K. (2012). A free energy principle for biological systems. *Entropy*, *14*(11), 2100–2121. https://doi.org/10.3390/e14112100

Friston, K. (2019). *A free energy principle for a particular physics*. https://arxiv.org/abs/1906.10184

Friston, K. J. (2011). Embodied inference: Or "I think therefore I am, if I am what I think." In W. Tschacher & C. Bergomi (Eds.), *The implications of embodiment: Cognition and communication* (pp. 89–125). Exeter: Imprint Academic.

Friston, K. J., Costa, L. D., & Parr, T. (2021). Some interesting observations on the free energy principle. *Entropy*, *23*(8), 1076. https://doi.org/10.3390/e23081076

Friston, K. J., Parr, T., Yufik, Y., Sajid, N., Price, C. J., & Holmes, E. (2020). Generative models, linguistic communication and active inference. *Neuroscience & Biobehavioral Reviews*, *118*, 42–64. https://doi.org/10.1016/j.neubiorev.2020.07.005

Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews*, *90*, 486–501. https://doi.org/10.1016/j.neubiorev.2018.04.004

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, *100*(1-3), 70–87. https://doi.org/10.1016/j.jphysparis.2006.10.001

Fuchs, T. (2017). *Ecology of the brain: The phenomenology and biology of the embodied mind*. Oxford University Press.

Gould, S. J. (2002). *The structure of evolutionary theory*. Harvard University Press.

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Hutto, D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. MIT Press.

Kirchhoff, M. D. (2018). Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese*, *195*(6), 2519–2540. https://doi.org/10.1007/s11229-016-1100-6

Kirchhoff, M. D., & Froese, T. (2017). Where there is life there is mind: In support of a strong life-mind continuity thesis. *Entropy*, *19*(4), 169. https://doi.org/10.3390/e19040169

Kirchhoff, M. D., & Kiverstein, J. (2021). How to determine the boundaries of the mind: A Markov blanket proposal. *Synthese*, *198*(5), 4791–4810. https://doi.org/10.1007/s11229-019-02370-y

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of The Royal Society Interface*, *15*(138), 20170792. https://doi.org/10.1098/rsif.2017.0792

Kiverstein, J. (2020). Free energy and the self: An ecological–enactive interpretation. *Topoi*, *39*(3), 559–574. https://doi.org/10.1007/s11245-018-9561-5

Korbak, T. (2021). Computational enactivism under the free energy principle. *Synthese*, *198*(3), 2743–2763. https://doi.org/10.1007/s11229-019-02243-4

Kostrubiec, V., Zanone, P.-G., Fuchs, A., & Kelso, J. A. S. (2012). Beyond the blank slate: Routes to learning new coordination patterns depend on the intrinsic dynamics of the learner—experimental evidence and theoretical model. *Frontiers in Human Neuroscience*, *6*, 222. https://doi.org/10.3389/fnhum.2012.00222

Kuhn, T. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Landecker, H. (2011). Food as exposure: Nutritional epigenetics and the new metabolism. *BioSocieties*, *6*(2), 167–194. https://doi.org/10.1057/biosoc.2011.1

Litwin, P., & Miłkowski, M. (2020). Unification by fiat: Arrested development of predictive processing. *Cognitive Science*, *44*(7), e12867. https://doi.org/10.1111/cogs.12867

Longo, G. (2018). How future depends on past and rare events in systems of life. *Foundations of Science*, *23*(3), 443–474. https://doi.org/10.1007/s10699-017-9535-x

Longo, G., & Montévil, M. (2014). *Perspectives on organisms: Biological time, symmetries and singularities.* Springer.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living.* D. Reidel Publishing Company.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge.* Shambhala.

Medaglia, J. D., Ramanathan, D. M., Venkatesan, U. M., & Hillary, F. G. (2011). The challenge of non-ergodicity in network neuroscience. *Network: Computation in Neural Systems*, *22*(1–4), 148–153. https://doi.org/10.3109/09638237.2011.639604

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M., & Nesselroade, J. R. (2014). New trends in the inductive use of relational developmental systems theory: Ergodicity, nonstationarity, and heterogeneity. In P. C. Molenaar, R. M. Lerner, & K. M. Newel (Eds.), *Handbook of developmental systems and methodology* (pp. 442–462). New York: Wiley.

Montévil, M. (2020). Historicity at the heart of biology. *Theory in Biosciences.* https://doi.org/10.1007/s12064-020-00320-8

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations.* New York, NY: John Wiley & Sons.

Orth, D., Kamp, J. van der, Memmert, D., & Savelsbergh, G. J. P. (2017). Creative motor actions as emerging from movement variability. *Frontiers in Psychology*, *8*, 1903. https://doi.org/10.3389/fpsyg.2017.01903

Oyama, S., Griffiths, P. E., & Gray, R. D. (Eds.). (2001). *Cycles of contingency: Developmental systems and evolution.* MIT Press.

Palacios, E. R., Razi, A., Parr, T., Kirchhoff, M., & Friston, K. (2020). On Markov blankets and hierarchical self-organisation. *Journal of Theoretical Biology*, *486*, 110089. https://doi.org/10.1016/j.jtbi.2019.110089

Parr, T., Costa, L. D., & Friston, K. (2019). Markov blankets, information geometry and stochastic thermodynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *378*(2164), 20190159. https://doi.org/10.1098/rsta.2019.0159

Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of Life Reviews*, *11*(3), 400–435. https://doi.org/10.1016/j.plrev.2014.03.005

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35. https://doi.org/10.1016/j.pneurobio.2015.09.001

Raja, V., Valluri, D., Baggs, E., Chemero, A., & Anderson, M. L. (2021). The Markov blanket trick: On the scope of the free energy principle and active inference. *Physics of Life Reviews*, *39*, 49–72. https://doi.org/10.1016/j.plrev.2021.09.001

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews*, *24*, 1–16. https://doi.org/10.1016/j.plrev.2017.09.001

Ramstead, M. J. D., Kirchhoff, M. D., Constant, A., & Friston, K. J. (2021). Multiscale integration: Beyond internalism and externalism. *Synthese*, *198*(S1), 41–70. https://doi.org/10.1007/s11229-019-02115-x

Rath, D., Amlinger, L., Rath, A., & Lundgren, M. (2015). The CRISPR-cas immune system: Biology, mechanisms and applications. *Biochimie*, *117*, 119–128. https://doi.org/10.1016/j.biochi.2015.03.025

Rose, L. T., Rouhani, P., & Fischer, K. W. (2013). The science of the individual. *Mind, Brain, and Education*, *7*(3), 152–158. https://doi.org/10.1111/mbe.12021

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, *5*(2), 97–118. https://doi.org/10.1080/17588928.2013.877880

Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: Building the web of life. *Nature Reviews Genetics*, *16*(8), 472–482. https://doi.org/10.1038/nrg3962

Speelman, C. P., & McGann, M. (2016). Editorial: Challenges to mean-based analysis in psychology: The contrast between individual people and general science. *Frontiers in Psychology*, *7*, 1234. https://doi.org/10.3389/fpsyg.2016.01234

Stewart, J., Gapenne, O., & Di Paolo, E. A. (Eds.). (2010). *Enaction: A new paradigm for cognitive science.* MIT Press.

Thelen, E. (2005). Dynamic systems theory and the complexity of change. *Psychoanalytic Dialogues*, *15*(2), 255–283. https://doi.org/10.1080/10481881509348831

Thelen, E., Corbetta, D., & Spencer, J. P. (1996). Development of reaching during the first year: Role of movement speed. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1059–1076. https://doi.org/10.1037/0096-1523.22.5.1059

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

Thompson, E. (2018). Review of Daniel D. Hutto and Erik Myin, Evolving enactivism: Basic minds meet content. *Notre Dame Philosophical Reviews*. https://ndpr.nd.edu/reviews/evolving-enactivism-basic-minds-meet-content/

Thompson, E., & Stapleton, M. (2009). Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi*, *28*(1), 23–30. https://doi.org/10.1007/s11245-008-9043-2

Turner, J. S. (2000). *The extended organism: The physiology of animal-built structures*. Harvard University Press.

van Es, T. (2021). Living models or life modelled? On the use of models in the free energy principle. *Adaptive Behavior*, *29*(3), 315–329. https://doi.org/10.1177/1059712320918678

van Es, T., & Kirchhoff, M. D. (2021). Between pebbles and organisms: Weaving autonomy into the Markov blanket. *Synthese*, *199*(3-4), 6623–6644. https://doi.org/10.1007/s11229-021-03084-w

Varela, F. J. (2011). Preface to the second edition of "De máquinas y seres vivos - Autopoiesis: La organización de lo vivo" by francisco varela. *Systems Research and Behavioral Science*, *28*(6), 601–617. https://doi.org/10.1002/sres.1122

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

Varela, F. J. (1991a). Making it concrete: Before, during and after breakdowns. *Revue Internationale de Psychopathologie*, *4*(6), 601–617. https://associationpsychanalytiquedefrance.org/publications/revues/n4-1991/

Varela, F. J. (1991b). Organism: A meshwork of selfless selves. In A. I. Tauber (Ed.), *Organism and the origins of self. Boston studies in the philosophy of science vol. 129* (pp. 79–107). Dordrecht: Springer. https://doi.org/10.1007/978-94-011-3406-4_5

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

Varela, F., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, *2*(4), 229–239. https://doi.org/10.1038/35067550

Veissière, S. P. L., Constant, A., Ramstead, M. J. D., Friston, K. J., & Kirmayer, L. J. (2019). Thinking through other minds: A variational approach to cognition and culture. *Behavioral and Brain Sciences*, *43*, E90. https://doi.org/10.1017/s0140525x19001213

Villalobos, M., & Ward, D. (2015). Living systems: Autonomy, autopoiesis and enaction. *Philosophy & Technology*, *28*(2), 225–239. https://doi.org/10.1007/s13347-014-0154-y

Vygotsky, L. (2012). *Thought and language. Revised and expanded edition*. MIT Press.

Waddington, C. H. (1957). *The strategy of the genes: A discussion of some aspects of theoretical biology*. George Allen; Unwin.

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321

Weron, A., Burnecki, K., Akin, E. J., Solé, L., Balcerek, M., Tamkun, M. M., & Krapf, D. (2017). Ergodicity breaking on the neuronal surface emerges from random switching between diffusive states. *Scientific Reports*, *7*(1), 5404. https://doi.org/10.1038/s41598-017-05911-y

Wiese, W., & Friston, K. J. (2021). Examining the continuity between life and mind: Is there a continuity between autopoietic intentionality and representationality? *Philosophies*, *6*(1), 18. https://doi.org/10.3390/philosophies6010018

Di Paolo, E. A., Thompson, E., & Beer, R. D. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. *Philosophy and the Mind Sciences*, *3*, 2. https://doi.org/10.33735/phimisci.2022.9187