A Moving Boundary, a Plastic Core: A Contribution to the Third Wave of Extended-Mind Research

Timotej Prosen • University of Ljubljana, Slovenia • timotej.prosen92/at/gmail.com

> Context • The current state of extended-mind research involves different frameworks, predictive processing and enactivism, among others. It is unclear to what degree these two frameworks converge toward a unified conception of the extended mind. **> Problem** • The third wave of extended-mind research expands the scope of what has been acknowledged as a legitimate case of extended mind under the parity principle and complementarity principle of the first two waves. The two central commitments of the third wave are: (a) That extended cognitive agents exhibit plasticity (b) that extended cognitive systems may not be organism-centered. I explore a general notion of boundary that might accommodate those two claims and provide a general criterion of what constitutes a case of extended mind. **> Method** • I employ the method of conceptual analysis. I explore several conceptions of mental boundary and plasticity with regard to the context of the wider conceptual frameworks within which they are embedded, namely predictive processing and enactivism. I confront the two frameworks with regard to how the notions they provide fare against the issues of third-wave extended-mind research. **> Results** • (a) I confront the notion of boundary of cognitive agents based on the Markov-blanket formalism with the enactivist notion of mental boundary based on operational closure and argue for the latter approach. (b) I argue that plasticity of mental boundaries exhibits two fundamental facets that can be distinguished and accounted for by recourse to Ashby's conception of ultrastability and the notion of perceptual inference as employed by the free-energy principle, if the two notions are integrated into the enactivist framework. **> Implications** • Predictive processing and enactivism cannot be reconciled regarding their respective notions of boundary of cognitive agents. In this regard, enactivism provides a better point of departure for the third wave of extended-mind research. Notions of active and perceptual inference based on the free-energy principle might nevertheless provide insights that enrich the enactivist position and lead to a more nuanced perspective on the extended mind. **> Constructivist content** • Constructivist epistemology forms the theoretical background of some key notions, utilized throughout the article, namely the conception of mind as autonomous and self-organizing.
> Key words • **Enactivism, extended mind, free-energy principle, Markov blanket, predictive processing.**

Introduction

« 1 » Where does the mind end and the rest of the world begin? Proposing the notion of *extended mind*, Andy Clark and David Chalmers famously argue that this boundary should not be sought at the edges of our skulls or even the skin of our biological bodies (Chalmers & Clark 1998: 7). Rather, they claim that the underpinnings of our minds can also include numerous external media, through which we have come to record, communicate and process our thoughts (ibid: 8). This conception of the mind has since gained much traction while also sparking ongoing polemics and debate.

« 2 » A fundamental point of dispute concerns the very notion of *boundary* between mind and world (Di Paolo 2009; Clark 2017). Here, I will examine which conception of boundary of cognitive agents is most appropriate for the extended-mind (EM) framework. I will approach the matter by relying on a convenient distinction between the three successive waves of EM research (Sutton 2010; Kirchhoff 2012) and align myself with the developments of the most recent third wave. Stated briefly, the third wave involves two crucial aspects. Firstly, the third wave attends to the *plasticity* of cognitive agents and the *temporally extended process* through which agents incorporate and are modified by external media (Kirchhoff 2012; Kirchhoff & Kiverstein 2020b). Secondly, the third wave advocates that the notion of extended mind should include relations that are *not organism-centered* – i.e., we need to take into account processes of distributed cognition and collective agency (Kirchhoff 2011; Kirchhoff & Kiverstein 2019; Loughlin 2020). Throughout the article, I will attempt to articulate a notion of boundary of cognitive agents, which could account for the two major commitments of the third-wave EM position.

« 3 » Such a notion of boundary can be approached from two distinct frameworks, namely *enactivism* and *predictive processing* (PP), both of which have recently been associated with third-wave EM. Some look to enactivism for notions of plasticity and dynamically constituted agents (Szanto 2013; Gallagher 2018; Kirchhoff & Kiverstein 2019). Objections arise however, that enactivism remains committed to the organism-centered view (Ryan & Schiavio 2019; Loughlin 2020). PP, on the other hand,

Handling Editor • Alexander Riegler • Vrije Universiteit Brussel, Belgium

seems to provide a more widely applicable conception of mental boundary via the notion of the Markov blanket (Kirchhoff et al. 2018; Kirchhoff & Kiverstein 2020a). The Markov-blanket approach seems more appropriate for describing cognitive systems that are not organism-centered (Kirchhoff et al. 2018; Ramstead et al. 2021). However, some express concern that the notion of the Markov blanket, taken by itself (Clark 2017), or even the PP as such (Di Paolo, Thompson & Beer 2022), cannot account for plastic transformations of cognitive agents and their boundaries. In line with these latter concerns, I will argue against the adoption of the Markov-blanket formalism, and advocate the claim that the enactivist conception of operational closure provides a better starting point for the third-wave EM approach. Subsequently, I will attempt to overcome the issue of organism-centeredness with recourse to PP and the related free-energy principle (FEP) in line with certain attempts to bridge the two frameworks (Clark 2016; Bruineberg, Kiverstein & Rietveld 2018; Gallagher & Allen 2018). From this vantage point, I will attempt to articulate an enactivist notion of plasticity, which might allow for the integration of individual organisms into collective and distributed forms of cognitive agency.

The three waves of extended mind

« 4 » The issues which I take as my point of departure are best approached by first outlining the third-wave position and by fleshing out its core commitments. So let me begin with a brief overview of the three successive waves of EM research. I shall present the three positions by recourse to the classic case of Otto and Inga proposed by Clark and Chalmers (1998). This thought experiment remains one of the most powerful illustrations of the basic commitments behind first-wave EM. Moreover, the issues some have taken with Clark and Chalmers's rendering of the Otto case will serve as a good entry point into the problems faced by EM research as it stands today. To give a quick recap of the thought experiment, Otto and Inga both set out to see an exhibition at a museum. Inga, who is mentally unimpaired, can simply recall the location of the museum in her head, along with all the necessary cues about how to find her way there. Otto, who suffers from Alzheimer's disease, must find a way to get around his inability to remember such details. So, he relies on a notebook, which allows him to keep track of his plans and guides him step by step to see them through, by providing a reliable and easily accessible source of details such as the location of the museum.

« 5 » What is being extended here, or so the argument goes, is Otto's mind, namely his dispositional beliefs about the exhibition and location of the museum, which are implemented not only by his brain but also by the notebook. The role of his mentally unimpaired counterpart Inga in the story is also quite straightforward – namely, it is to demonstrate that the same mental "coarse-grained" function is being implemented in two different ways. Or this, at least, is how the first wave of EM would frame the argument. Indeed, as soon as we begin to unpack this example, the issues that divide the different waves of EM research become manifest.

« 6 » Clark and Chalmers propose a principled way of determining what counts as a legitimate case of EM. Their criterion has been dubbed the *parity principle* (Sutton 2010: 193) and is considered to be the hallmark of the first-wave position. They propose:

> "If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time), part of the cognitive process." (Chalmers & Clark 2008: 8)

« 7 » The parity principle has, however, been widely disputed (O'Brien 1998; Menary 2010; Sutton 2010). The main point of contention has been the physiological comparability (O'Brien 1998), or even structural parity (Menary 2010; Sutton 2010; Kirchhoff 2012; Kirchhoff & Kiverstein 2019), of the extended mental functions and their biological counterparts. With regard to Otto's case, it has been argued that our technical information storage devices are only superficially comparable to our biological capacities for recollection. It is claimed, for example, that the main difference is that human biological memory is non-symbolic, holistic and highly context-dependent (O'Brien 1998). These kinds of issues were tackled by the second wave of EM research (Menary 2010; Sutton 2010), which disposes of the criterium of parity in favor of a principle that encapsulates the creative and non-anthropomorphic character of extended mental processes. The so-called *complementarity principle*, proposed by John Sutton, maintains that internal and external processes play complementary roles in constituting new mental functions, ones which might carry no resemblance to those we are accustomed to (Sutton 2010: 194). For example, it is not despite but because of functional disparities that our biological memories are supplemented by more stable external scaffolding (ibid: 205). According to the second wave, extension therefore implies functional transformation on our part (Menary & Kirchhoff 2014: 5), and it is primarily because of the advantages of external media that we extend our mental capacities in the first place.

« 8 » The third wave can be said to depart even further from the first wave on whether extended cognitive systems need to mirror predefined internal structures (Loughlin 2020: 6). Firstly, it attends to capacities for *plasticity* (Kirchhoff & Kiverstein 2019; Loughlin 2020), which allow cognitive agents to transform in order to incorporate and integrate with aspects of their environments. The extended cognitive agent is said to take shape through a temporal process whereby the agent is continually modified by external media it incorporates (Kirchhoff 2012; Ryan & Schiavio 2019). Such radically plastic agents are not simply complemented by external media. Rather, they constitute mutable "loci of coordination" (Sutton 2010: 213), whose individual identity is open to constant renegotiation. The second distinctive feature of third-wave EM is the claim that such loci of coordination *need not be organism-centered* (Loughlin 2020; Kirchhoff 2012; Kirchhoff & Kiverstein 2019). The third wave rejects the view that the extended cognitive system should necessarily be assembled and organized by individual organisms (Kirchhoff & Kiverstein 2019: 18). Individuals might, for example, integrate with collective and institutional practices (Szanto

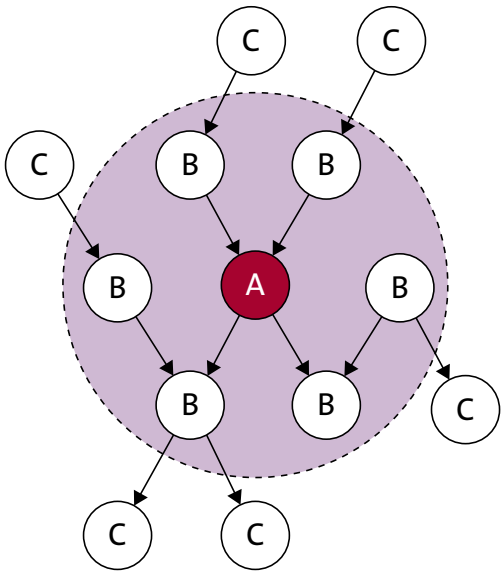

Figure 1 • A simple schema of a Markov blanket, with internal states (A), blanket states (B), and external states (C).

2013), which as a whole exhibit features of cognitive agency, without any particular individual being in the position of control and coordination over the larger whole (Kirchhoff & Kiverstein 2019: 19).

« 9 » The radical view of plasticity and the socially distributed nature of cognitive agents that the third wave advocates raises doubts however, whether such agents may be conceived of as bounded entities at all (Loughlin 2020: 7). If not, then the third-wave position seems to depart from EM altogether, because it seems to be lacking any criterion of how to demarcate an extended agent (ibid). However, as I will argue in the following sections, a distinctive third-wave conception of boundaries of mental agents can be formulated. In order to be compatible with the two central claims of third-wave EM, such a notion must satisfy two conditions. Firstly, the boundaries of a cognitive agent must be shown to be plastic – an account must be given as to how the agent reshapes its boundaries over time. And secondly, such boundaries must not be necessarily centered around an individual organism.

« 10 » One such conception of boundaries has already been proposed and explicitly related to the third-wave approach, namely the Markov-blanket formalism (Kirchhoff & Kiverstein 2019: 62). I shall proceed by discussing this notion in order to determine whether it does justice to the two basic requirements of the third-wave approach.

Markov blankets of extended minds

« 11 » The idea that boundaries of a cognitive agent might be delineated via the Markov-blanket formalism is tightly interwoven with the overarching account of cognition provided by PP. So, before the Markov-blanket formalism can be discussed, a brief overview of the PP framework is in order. According to PP, the mind is in the business of making predictions about incoming sensory inputs and their causes (Hohwy 2013: 14), and actively counteracting the discrepancy between said predictions and given sensorial experience (ibid: 77). The mismatch between the given sensorial experience and what the mind predicted it to be is what drives both learning and behavior, which both ultimately serve to satisfy the same imperative – to minimize this mismatch (prediction error). The main difference between these two basic mental functions lies only in the way prediction error is minimized; while perceptual learning updates the predictions (priors) to fit new sensory data, behavior, conversely, strives to change, or move around the world itself, so that the sensorial experience would fall in line with the predictions (ibid: 43).

« 12 » PP requires some exact way of characterizing the discrepancy between the agent's predictions and its sensory states, and how this discrepancy can be minimized by acting on the external causes of sensory experience. This is done by distinguishing between the internal states of the cognitive agent and the states of the external environment, by the intermediary set of the agent's sensory and active states, through which all agent-environment interaction is said to take place (Parr, Pezzulo & Friston 2022: 44). The *Markov blanket* has been proposed as a way to formally capture this set of distinctions (Friston 2013; Hohwy 2016, 2017).

« 13 » The Markov-blanket formalism is a notion taken from statistics, which fits nicely with the core aspects of the predictive view of the mind (Pearl 1988). The Markov blanket is a set of states, or rather variables, describing those states, that statistically insulate internal states of a system from external states. An internal state, or a variable, (as indicated by "A" in Figure 1) is statistically insulated from external state C by blanket state B, when, given that we know the value of B, the value of C provides no further basis for inference about the value of A, and vice versa. We can also frame these relations in terms of causal interactions of a dynamical system (Parr, Pezzulo & Friston 2022: 43), i.e., by saying that any causal influence between internal and external states is mediated by the Markov blanket. In other words, external states can influence internal states only indirectly by first modifying the blanket states, which in turn modify internal states, and vice versa.

« 14 » The basic idea behind the Markov blanket as a mental boundary is straightforward. The external states (C) are said to denote the states of the external environment, internal states (A) stand for the internal states of the agent, and the blanket states (B) denote all the states of the agent that mediate the agent's sensorimotor interaction with the environment (ibid: 43). The causal insulation, which our bodily boundaries provide

to our cognitive apparatus, also acts as a kind of epistemic insulation for our minds. Suppose that you are enveloped inside a Markov blanket. You have no direct knowledge of the states of your environment, just as those states do not affect you directly, but only through the intermediary of your boundary states. What you know about the world is mediated by the interchanging states of your Markov blanket, particularly the states of your sense organs. The predictions you make about future states of your sensory states can be said to be about the world on some level. This is because they are (statistically) indicative about the sequence of events in the world and can be said to model the states of the world. The blanket states thus play the role of evidence, which simultaneously point to, and seclude the agent from, the hidden states of the external world.

« 15 » Where then might such a boundary be located? This question has given rise to much controversy. The Markov-blanket formalism has been invoked by both internalist (Hohwy 2016, 2017, 2019) and externalist (Clark 2017; Kirchhoff & Kiverstein 2019; Ramstead et al. 2021) arguments. The former argues for a brain-bound conception of the mind (Hohwy 2017) that places the Markov blanket of a cognitive agent at the sensory and motor junction between the nervous system and the rest of the organism (Hohwy 2019: 12). The latter authors offer externalist alternatives, which apply the Markov-blanket formalism beyond the level of the nervous system and individual organisms.

« 16 » To understand the reasoning behind some internalist and externalist positions, let us consider two diverging takes on the case of Otto and his notebook. Jakob Hohwy argues that our minds cannot extend to include such props, because they seem to fall outside of the boundary of the Markov blanket enveloping the agent's nervous system:

> "An agent can grasp and use her phone [or notebook] only because she has a more or less precise and accurate internal representation of the phone, the things in her drawer may occlude it, and the causal interactions between her fingers, eyes and ears, voice and the states of the phone." (Hohwy 2016: 270)

« 17 » Michael Kirchhoff and Julian Kiverstein, however, argue for a very different interpretation of the same example:

> "When Otto forms the intention to visit Modern Museum of Art (MoMa), he predicts the sensory states associated with the action of heading off to the museum. He then acts in such a way as to fulfil his predictions, thereby reducing prediction error. Among the actions Otto must perform to fulfil his predictions is the consultation of his notebook." (Kirchhoff & Kiverstein 2019: 78)

Because the notebook is an essential component of predictive-error minimization for some of Otto's mental tasks, the authors count it as part of the Markov blanket that defines Otto as a cognitive agent (ibid).

« 18 » The Markov-blanket formalism thus lends itself to both internalist and EM positions. The formalism is applied to the case of Otto and his notebook in two radically different ways. Both applications can appear justified, however, depending on the way we frame the Otto case and which cognitive function we deem to be centrally relevant in this case. When we consider only the interaction between Otto and his notebook, the notebook seems to fall on the side of the external environment. However, if we were to shift our attention to Otto's task of navigating his way to the museum, his notebook plays the role of a constituent of his mind, as Otto clearly could not effectively represent the museum without relying on the notebook as evidence.

« 19 » In line with such considerations some have argued that the internalist and externalist applications of the Markov-blanket formalism are both justified to some degree (Ramstead et al. 2021: 42). This conclusion, however, they claim, does not falsify the EM position (ibid.). Markov blankets allow for nesting and simultaneous overlap, a feature that some have ascribed to boundaries of cognitive agents as well (Clark 2017; Allen & Friston 2018; Kirchhoff et al. 2018; Ramstead et al. 2021). On this view, cognitive agents can be said to be embedded and overlapping across various scales, ranging from cells to individual organisms and up to collective cognitive systems involving environmental niches and cultural practices (Ramstead et al. 2021: 43). Such a multiscale conception of mental boundaries opens the possibility of reconciliation between internalism and externalism, because it is perfectly compatible with Hohwy's neurocentric application of the Markov-blanket formalism, while at the same time allowing for the claim that the neural agent forms a constituent part of some higher-order agent, say, Otto interacting with his notebook.

« 20 » I claim however, that such an account by itself does not fully explain instances of extended mind such as the case of Otto. The notion of nested Markov blankets seems appropriate to describe entities that constitute a higher-order system *at a given moment*, such as cells composing a multicellular organism. The notion of extended mind, on the other hand, is intended to capture cases where cognitive agents actively incorporate aspects of their external environments *over time* and whereby such agents are themselves modified by such a process. A satisfactory third-wave EM position must account for the *plasticity* of the boundaries of cognitive agents, i.e., how their boundaries transform over time. We have yet to account for how someone like Otto, who must have learned to rely on his notebook at some point in his life, is able to incorporate external media so as to gradually constitute an extended cognitive system.

« 21 » Although some externalist proponents have sought to combine the multileveled conception of mental boundaries with an account of plasticity of Markov blankets (Clark 2017; Kirchhoff & Kiverstein 2019, 2020a; Ramstead et al. 2021), the two approaches remain importantly distinct. As Clark (2017: 11) points out, an account of plasticity of Markov blankets must adopt a process ontology of cognitive agents, which does not identify individual persisting agents with any specific stationary Markov blanket at all. A cognitive agent exhibiting plasticity on the level of its Markov blankets cannot be fully delineated by any specific configuration of states, because such an agent must be able to continually build and rebuild its Markov blankets without thereby dissolving its individual identity (ibid). One might of course still argue that the Markov-blanket formalism remains appropriate for describing the boundaries of a plastic temporally evolving agent *at a given moment in time*. However, an issue arises whenever we are dealing with a complex cognitive system

that allows for multiple equally valid Markov-blanket delineations, such as Otto and his notebook. In such cases, the question remains as to which Markov-blanket configuration can be taken to pertain to the boundary of some particular temporally evolving agent. A boundary configuration therefore cannot always be ascribed to an agent on the sole ground of exhibiting characteristics of a Markov blanket. Rather, as Kirchhoff and Kiverstein (2019: 73) argue, we can pick out a Markov blanket pertaining to a temporally extended agent only by recourse to an account of how that boundary was generated over time.

« 22 » The Markov-blanket formalism, taken by itself, cannot provide a reliable method of demarcation of boundaries of temporally evolving agents. It seems that any notion of such moving boundaries must be developed by recourse to an explicit account of how they are generated over time. To this end many look to enactivist notions such as *autonomy* and *operational closure*, as they seem to offer both an account of how such boundaries are actively generated by a temporally enduring agent (Clark 2017; Kirchhoff et al. 2018; Ramstead et al. 2021) and how such boundaries evolve and change shape over time (Di Paolo 2009; Di Paolo, Thompson & Beer 2022). I shall follow their cue in the subsequent section.

An enacted boundary

« 23 » In exploring the approach to the boundaries of cognitive systems, I shall take the interrelated notions of operational closure and autonomy as my point of departure. *Operational closure*, as defined by Francisco Varela (1981: 15), is the distinctive characteristic of *autonomous systems*, which (a) through their interaction continually generate the network of relations characteristic of the system and (b) constitute the system as a unity, distinct from its surroundings. An autonomous system is not closed off from interactions with an environment, however. Rather, the system must actively counteract outside perturbations and continuously maintain such interactions with the environment that allow for its continued integrity (Varela 1997: 79). The two features of operational closure are thus interrelated. An autonomous system is distinguished from an environment by the very process that continually generates and maintains the identity of the enduring whole (Varela 1981: 16).

« 24 » The notion of operational closure thus seems well suited to explain how an agent actively distinguishes itself from an environment over time. Claiming that autonomous systems maintain their identity may raise doubts about whether the enactivist perspective is apt to accommodate the capacity for plasticity of temporally evolving agents. However, the identity of autonomous cognitive agents might not be as predefined and rigid and as their shorthand definition may imply. To see why this is the case, we need to further elaborate what makes an autonomous agent cognitive. Such agents can be ascribed cognition, because they are able to interact with the environment in such a way as to retain their organization (Varela 1992: 255). However, one may inquire further into just how such cognitive agents come to know which kinds of interactions are beneficial for their self-production. As Ezequiel Di Paolo (2009: 15) points out, the ability to maintain favorable exchanges with the environment is not explained by the concept of autonomy itself. A bacillus is autonomous until it runs out of nutrients or is dissolved by too much heat, but its autonomous nature does not explain why it should actively seek out food or avoid dangerous situations. The problem is that autonomy is a holistic and all-or-nothing quality of the agent, which does not enable us to capture the ways in which an autonomous agent can change over time to acquire new modes of behavior. The bacillus could be said to maintain its autonomy until the day it dies and ceases to self-generate altogether. I shall therefore, following Di Paolo, supplement the notion of autonomy with an account of *adaptivity*. The latter notion stands for the capacity of an agent to regulate itself with respect to the bounds of its viability (Di Paolo 2005: 430). An account of adaptivity offers insights into how an organism such as a bacillus comes to "know" when it is approaching the limits of its ability to sustain itself, and how it modifies its behavior accordingly. However, more importantly for my purposes, the notion of adaptivity might enable us to discern a degree of plasticity in autonomous agents (Di Paolo 2009: 18) by accounting for the transformations that such agents undergo precisely in order to remain viable as a systemic whole.

« 25 » An account of adaptivity can already be found in the seminal cybernetic work of Ross Ashby (1960), who proposed some rudimentary principles of such adaptive capabilities. Ashby speculated that a cognitive agent must ultimately be based on two essential components. The first couples the agent with its environment in an immediate way, such that an external sensory stimulus elicits a specific motor response. The second component tracks how the agent stands with regard to its own essential variables, such as homeostatic limits that an organism must maintain to stay alive. The second subsystem thus perpetuates autonomous organization by monitoring and maintaining essential structures. At the same time, the essential variables, in turn, modulate the "cognitive" principles of sensorimotor coordination. When the essential variables go out of acceptable bounds, the sensorimotor system is compelled to permutate (randomly or otherwise), until the agent "finds" a new mode of sensorimotor interaction with the environment, which returns the system's essential variables to within viable bounds.

« 26 » In Figure 2, the functional (input-output) boundary of the agent is explicitly marked. However, Ashby's model implies another important distinction that sets apart two aspects of the agent itself. The essential variables (E), on the one hand, must remain unchanged[1] as long as the agent can be said to maintain its identity. The sensorimotor system (R), on the other hand, exhibits plasticity. Moreover, the latter system may be said to undergo transformations precisely so as to maintain the essential components unaltered. I propose, therefore, that we construe the sensorimotor system (as defined by Ashby) as the *boundary of the autonomous cognitive agent*. Because the sensorimotor system is set up so as to continually maintain the essential variables within viable bounds, it might be thought of as a boundary that insulates the central structures from environmental perturbations. The spatial connotations of boundary here are mostly

1| Or rather, maintained within viable bounds, to be more precise.

metaphorical, however: on this account, the boundary is not *topological* but rather *operational*. Unlike the Markov blanket, the sensorimotor system does not insulate the interior essential variables from direct causal influences; rather, what it does is counteract those influences over time, so as to maintain a relatively stable interior structure.

« 27 » Although Ashby did not develop his conception of adaptivity in the direction which might be explicitly related to EM, I claim that his notions are perfectly suited to characterize the moving boundary of extended cognitive agents. Because the range of plasticity of the sensorimotor system is constrained only by its manner of coupling to the agents' essential variables, the operational boundary of the autonomous agent can come to incorporate all manner of external media. Such a notion of boundary might therefore speak to a number of cases of mental extension. Particularly striking examples are manifested in cases of pathology of some cognitive trait, such as reports of blind people, who, after an initial period of habituation of maneuvering with a white cane no longer take notice of their visual prosthetic, but sense the obstacles directly through it, as if it had become an extension of their bodies, susceptible to the sensation of touch (Merleau-Ponty 2005: 174). However, we may find such sensorimotor plasticity at work to some degree almost anywhere we look – from the way that driving a car or playing a video game may temporarily alter the sensorimotor structure of our consciousness (Clark 2003: 120) to how the mode of recollection and planning is structured by usage of something like a notebook (Clark 2008; Di Paolo, Cuffari & De Jaegher 2018) or the way our sense of time and place is altered by smart devices (Clark 2003: 43).

« 28 » How the outlined enactivist notion of mental boundary relates to specific instances of extended mind could be explored in much further detail. What I want to conclude with here is rather a question of whether such a notion of boundary speaks to the two challenges laid out by third-wave EM, namely (a) to account for cases of EM that are not organism-centered and (b) to account for the temporally unfolding plasticity of the cognitive boundary. With respect to (b), I claim that the enactivist approach overcomes the issues concerning the Markov-blanket formalism. Notions of autonomy and adaptivity provide a way of distinguishing between essential enduring aspects of cognitive agents and a boundary that insulates the latter from external perturbations over time. Moreover, Ashby's construal of the relation between the sensorimotor system and essential variables provides an account of how such operational boundaries are shaped and reshaped over time. I claim therefore that the enactivist approach does justice to challenge (b) by articulating a convincing account of boundaries that may move and change shape over time.

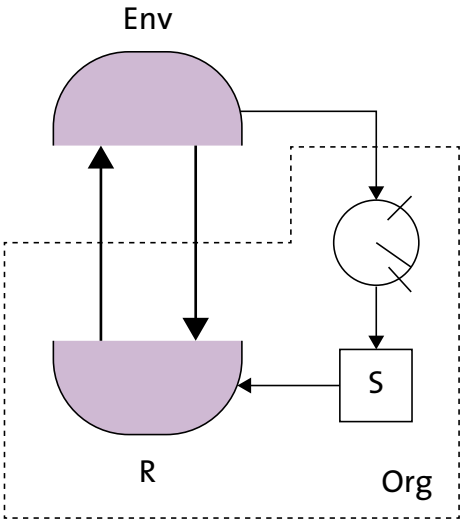

Figure 2 • Ashby's schema of an ultrastable system. The agent (Org) is circumscribed by the dotted line. The interaction with environment (Env) takes place on two levels. On one level, the environment may act on the sensorimotor system (R) and elicit some behavioral response or vice versa, i.e., the agent may modify the environment and elicit some sensory feedback. The environment may also cause the essential variables (represented by the meter) to go out of bounds. If the perturbation of essential variables is not adequately counteracted by the sensorimotor system (R), (S) introduces random changes to R until they are able to bring the essential variables within bounds.

« 29 » What about challenge (a) – must such moving boundaries necessarily be centered around individual organisms? Undoubtedly, enactivism casts a specific light on cognitive agents: one that construes such agents as essentially living individuals, i.e., individuals interested in their continued existence, and whose very tendency toward maintaining some particular mode of identity provides a stable foothold for the expansion of their boundaries. The enactivist view of cognitive boundaries, as I have sketched it here, is therefore committed to a certain organism-centeredness. It remains a perspective wherein the individual organism remains central to the organization of extended cognitive systems and wherein the mind incorporates alien features of its surroundings ultimately to subdue them to its particular interests, to bind and integrate them with a pre-existing "core" of their individual structure. Nevertheless, I claim that such a notion of extended mind can be fruitfully confronted with attempts to break out of organism-centeredness. Let me point out two such issues.

« 30 » For one, an important challenge to the organism-centered view is to be found in examples of institutions and collectives that display intelligent behavior, without any single individual being in a position of direct control over the larger whole (Kirchhoff 2012). Such higher-order cognitive systems can be considered a form of extended mind

if we can account for how individuals accommodate the larger social whole through their internal structure. This causes issues for any notion that allows for plasticity only on the level of sensorimotor means of maintaining a predetermined and fixed individual identity. There have been attempts to tackle this issue from the standpoint of enactivism under the heading of *participatory sense-making*. I am very sympathetic to the ideas developed in this context, because they indicate a general direction of an enactivist notion of EM, which I will attempt to flesh out in the subsequent sections. One such claim is that the social realm should be understood neither as a sum of predefined individuals, nor on the level of some higher-order collective agent, but in the intermediary domain between the two levels, i.e., in the way individuals and the emergent social realm mutually shape one another (Di Paolo & De Jaegher 2017: 99). Another particularly interesting claim is that such social relations involve a decentering of individual agents (Fuchs & De Jaegher 2009: 476), and that such decentering is achieved by agents that mutually modulate their autonomous self-organization (De Jaegher & Di Paolo 2007). This suggests that their notion of center, which is said to be displaced, can therefore be cast in terms of my distinction between the central aspects of autonomous identity and their operational boundary. In line with their approach, I am therefore motivated to consider a notion of autonomous agency, which not only possesses a moving boundary, but also a *plastic core* of individual identity.

« 31 » There is another issue with the organism-centered view, which I suspect is closely related to the claims about participatory sense-making – not every significant agent-environment coupling seems beneficial to the viability of any individual agent. We are constantly engaged with artefacts of aesthetic, religious or other kinds of affective importance, even though such objects do not seem necessarily beneficial to us in particular. Their lack of obvious utility is in stark contrast to the intense reverence with which they are treated; objects such as flags, religious artefacts, works of art or personal trinkets often receive greater attention and care than our immediate means of maintaining our individual viability. Now, there are various ways of explaining such seemingly irrational attachments, and the notion of the extended mind certainly is not the most obvious choice. However, it is precisely the enactivist way of framing relations of utility that provides us with a vantage point for such an attempt. Since use value is made sense of in relation to a central domain of autonomously self-maintaining aspects, such objects might appear to be without utility precisely because they externalize those very central aspects of the agent's identity, that is to say, because engaging with such objects alters the very parameters of the agent's identity to be maintained. This issue challenges the organism-centered enactivist position in much the same way as the claims of the participatory sense-making approach. Both point beyond the cognitive architecture of a *moving boundary encircling a stable central core* and toward a *deepened notion of plasticity*, which I shall pursue in the following section.

Bounds of plasticity

« 32 » The remaining issues of the enactivist EM position converge and point toward a further development of our understanding of the extended mind. They seem to necessitate a radicalized notion of plasticity, such that it covers not only the operational boundary, but also the "central" core of autonomous identity. In other words, plasticity must be found not only on the level of sensorimotor ways of maintaining the viability of essential structures, but also on the level of the essential features themselves, i.e., the defining parameters of autonomous identity.

« 33 » Although these claims seem to run counter to the basic enactivist principles of autonomy, the divergence is subtle. My aim is to argue for the notion of an autonomous agent that is not only able to counteract perturbations of its essential features but is also able to redefine which states count as essential and which perturbations are to be counteracted. The framework as outlined above allows for a simple step in this direction: we can accept that autonomous cognitive agents are organized by bounds of viability set to maintain those very structures; however, these bounds need not remain fixed or even imply the sole viability of a single individual agent. Hence, I would argue for the notion of a *moving boundary encircling a plastic central core*.

« 34 » In the remainder of this discussion, I will try to unpack this general conclusion. One way to do this is to return to some points of PP that converge on the enactivist framework, developed by a closely related approach called *the free-energy principle* (FEP). FEP takes up the central idea of PP, namely that the two fundamental features of the mind are predicting sensorial experience and actively minimizing prediction error (Parr, Pezzulo & Friston 2022: 198). In contrast to PP, FEP starts from the question of how living organisms persist and act on the world (ibid: 7). FEP thus develops the predictive approach to the mind with a specific interest in how living agents self-organize and actively engage with their environments, which is why many have likened FEP to enactivism (Kirchhoff 2018; Gallagher & Allen 2018; Ramstead, Badcock & Friston 2018; Ramstead et al. 2021). After a brief overview of FEP, I shall point out a possible convergence between the two approaches and also a distinctive feature of FEP, which is relevant to third-wave EM.

« 35 » As the very name suggests, the conceptual framework of FEP is based on the notion of variational free energy (FE). FE is what is said to underlie the behavior of cognitive agents and integrate the various mental functions under a common goal, namely, the goal of minimizing this quantity called "variational free energy." So, what precisely is this quantity? On the one hand, FE can be thought of as a mathematical notion, which (to put it simply) is closely related to the concept of entropy. To minimize the free energy of a physical system (such as an organism within its environment) entails that the entropy of that system is maintained under some limit (Friston 2013: 5). From this perspective, minimization of FE describes *action*, by which an agent maintains a low-entropy state and thus ensures its own survival (ibid). On the other hand, FE can be understood in information-theoretical terms. On this level, FE can be interpreted as a measure of mismatch between sensorial experience and mental predictions or models of sensory data (Friston 2009: 293). The more that sensory data deviates from the expected course of events, the greater the

amount of free energy. Minimization of FE can therefore also be achieved by way of perception and learning, i.e., by adapting our predictions to given sensorial experience.

« 36 » Similarly to enactivism, FEP thus interrelates the self-organizational (or entropy-defying) and the cognitive aspect of mental functioning. The basic idea, and the most attractive feature of FEP, is that the imperative to minimize free energy provides an elegant unifying structure to how we learn about and how we respond to the environment, one that conceives of both perception and action as two facets of the same adaptive strategy.

« 37 » In the case of a mismatch between prediction and sensorial experience, it is natural to assume that an agent capable of learning about its environment should update its predictions so as to accommodate the surprising experience. And indeed, this is said to be one way to minimize free energy, termed *perceptual inference* (Friston & Stephan 2007: 11). Taken as such, however, it does not yet constitute an adaptive response. An adaptive agent need not only predict what happens, but also act on the world to its advantage. FEP thus makes the additional assumption that the predictive model of an adaptive agent is such as to imply the continued existence of that very agent, i.e., it describes the course of agent–environment interactions that are viable for that agent[2] (Parr, Pezzulo & Friston 2022: 46). This assumption implies that what is surprising (with respect to the prediction) is also entropic and detrimental to the agent (ibid: 2). Therefore, action can also be framed as active avoidance of surprisal, i.e., deviation from what is taken to be the most likely course of events. This way of dealing with prediction error is called *active inference* (Friston 2009: 299), because the action on the environment is controlled by a predictive model, ultimately by the same model that is constrained by *perceptual inference*. Perceptual and active inference thus form the two fundamental aspects of the same cognitive architecture – one constrains the prediction by changing it to fit perceptual data, while the other modifies the environment to fit the prediction. Both minimize the same mismatch, but by proceeding in different although complementary and mutually supporting ways.

« 38 » The elegant simplicity of such a conception of the mind can be deceptive, and it certainly does not come without some controversy. A major concern is whether such an emphasis on active inference undermines the representationalist notion that the agent's predictions model the objective state of its environment (Gallagher & Allen 2018: 2634). The prevailing interpretation of predictive models is that such a representation of a pre-given environment is indeed the job of perceptual inference – its function is said to be to adequately model the states of affairs in the world (Friston & Stephan 2007: 15; Hohwy 2013: 52). However, as I have pointed out, a predictive model as understood by FEP is only as good as the active inference that supports it; its main purpose is to guide adaptive action. Elsewhere, Friston (2013: 8) admits to a fundamental ambiguity of whether internal states predict external causes of their sensory states or actively cause them through action.

« 39 » However insignificant the agent's contribution to the causal sequence of events may be, it is therefore somewhat misleading to speak of prediction of external causes at all.[3] This diction may obscure the causal role that mental contents have to play. If the agent does indeed actively strive to minimize prediction error, then its mental contents must ultimately be about its preferred internal states and situations, not some pre-existing state of affairs. This is the crucial insight behind the notion of active inference, namely that a predictive model may function as a kind of self-fulfilling prophecy (Clark 2018: 524). This suggests that the agent cannot be said to represent the environment as such; its predictive model functions more like a premonition of what the agent might become and a demand on how the environment should be. The agent operates not on an objective model of an environment, but on a kind of blueprint of itself and its own requirements. Or, to use Ashby's terminology, it could be argued that the agent models its own bounds of viability.[4]

« 40 » My claims are closely in line with a number of recent takes on related issues, which draw close connections between the notion of active inference in FEP and enactivism (Kirchhoff 2018; Gallagher & Allen 2018; Ramstead, Badcock & Friston 2018; Ramstead et al. 2021). As proponents of such approaches point out, active inference implies a conception of autonomous agency (Ramstead et al. 2021: 53) and the claim that cognition is fundamentally related to adaptive self-regulation (Kirchhoff 2018: 2538). The notion of active inference can therefore be brought closely in line with the enactivist approach as outlined in the previous section. However, the crucial thing I want to highlight here is the significance of the notion of *perceptual inference* for the further development of the enactivist position. As was mentioned, perceptual inference is said to involve the updating of prior beliefs or expectations to better fit new experiential data. Now, from the perspective of FEP, which is closely in line with that of enactivism, such prior beliefs and expectations are best taken not as modelling pre-existing environmental features, but as tracking and co-constituting the agent-world relations that are viable for that agent. However, what then are we to make of the process of updating such "beliefs" or learning through perceptual inference? If we keep in mind that such beliefs also serve as a basis for active inference through which the agent acts on the world and regulates itself, then a modification of any such model clearly implies the alteration of this very agent. Perceptual inference therefore implies a certain *plasticity* of adaptive agents. One should also note that the kind of plasticity is particularly profound – in enactivist terminology it could be said that it involves not only the alteration of manners by which agents counteract given

2| Karl Friston (2013: 5) formalizes this assumption in terms of the agent and environment composing a stable and ergodic system.

3| Except in a narrow sense of being statistically correlated with external causes, which does not preclude that those causes be themselves set in place by internal mechanisms of active inference.

4| Anil Seth (2015: 9) arrives at similar conclusions but argues that active inference is therefore better fitted to explain homeostatic control, then prediction of exteroceptive data. The most novel and exciting aspect of FEP, though, is perhaps precisely in the prospect of a single unifying account of interoceptive, proprioceptive and exteroceptive processing.

perturbations but also affects the parameters that define what counts as perturbation in the first place. Another way of putting it would be to say that perceptual inference involves *plasticity on the level of bounds of viability*.

« 41 » Independently of whether Ashby's notion of bounds of viability in its strict technical sense is applicable in the context of FEP, it could point toward a possible convergence between the two frameworks I have been working with. On the one hand, as I have argued in the previous section, Ashby's terminology allows us to further develop some of the enactivist principles of autonomy and self-organization in such a way as to allow for a certain degree of plasticity of cognitive agents. On the other hand, as I have argued in this section, the same terminology may be employed in the context of FEP to indicate another aspect of plasticity of such self-organizing agents, i.e., that perceptual inference as interpreted through the enactivist-leaning lens of FEP may involve plasticity on the levels of bounds of viability. At the very least, this notion therefore indicates the lines along which PP and FEP might significantly contribute to the enactivist position, especially regarding the notion of extended mind.

A unified third-wave perspective?

« 42 » As I have tried to show in the previous section, we can interpret some of PP's findings through the lens of enactivism, as does FEP (to some extent). This move might considerably enrich our understanding of EM, and add a new dimension to the enactivist conception of a *moving boundary*, namely a *plastic core* of individual identity. Such a perspective has the potential do justice to both basic tenets of the third-wave approach, namely (a) the commitment to *plastic boundaries* and (b) the claim that cognitive agents may *not be organism-centered*, for the following reasons. The first requirement was already satisfied by the enactivist notion of an operational boundary. The concern that the enactivist position remains organism-centered was addressed by the notion of a plastic core as developed through the lens of FEP. It is certainly not immediately obvious how such a radicalized notion of plasticity of individual identity allows agents to integrate with collective institutions and practices. I believe however, that the notion of a plastic core allows for steps in this direction because it establishes that an extended cognitive agent need not be organism-centered since it does not possess a predetermined center of individual identity at all, organismic or otherwise.

« 43 » So where are we to look for instances of extended mind involving a plastic core? To stick to the familiar story, one need only consider where Otto is heading, namely, to visit MoMa to see an art exhibit. A *work of art* might constitute just such a kind of mental extension. The enigmatic lack of obvious functionality made artistic objects rather unattractive candidates for the traditional perspective on the extended mind. Yet the way aesthetic objects elude pragmatic utility may turn out to be closely in line with the way I propose to move beyond the organism-centered dimension of the extended mind.

« 44 » Consider two aspects, which I find prevalent in aesthetic experience. For one, an aesthetic object presents itself as peculiarly salient and valuable, which is manifested both in the way it attracts our attention and in that we feel the need to preserve its aesthetic properties from degradation. On the other hand, these same properties can serve as (relatively) stable footholds, in relation to which our sensorimotor capacities organize into distinctive patterns. This is reflected, for instance, in how certain works of art inspire the creation of other works, thus standing at a point of origin for a distinct style of artistic expression. The influence such objects exert may, of course, go beyond overt artistic expression to reconfigure the general manner of their spectator's view of the world, both in terms of how they perceive it and how they strive to change it. In this regard, the work of art functions as a kind of "catalyst" for the formation of a certain aesthetic style of seeing and of actively structuring the world.

« 45 » Such characteristics display an interesting congruence with the perspectives on cognition developed in the previous section. Specifically, it leads me to consider a work of art as an *externalized bound of our viability*. Its affective salience may be compared with the valence of "physical" pain or pleasure, usually associated with the disturbance of our homeostatic bounds of viability. As such, aesthetic experiences also exert a similar motivating force, which drives and sculpts behavior. What is peculiar about aesthetic experience, and indeed about many other kinds of affective responses, is that they give rise to behavior that does not necessarily seem aimed toward the well-being or viability of the individual agent, or at least not exclusively toward the preservation of the agent alone, but also some larger part of its environment. Already on this level one might speak of a certain extendedness of bounds of viability, as what they track expands beyond the individual agent's body. Crucially, however, a phenomenon such as a work of art may be captured by the notion of extended *mind*, because it is also involved in the process by which the bounds of viability are (re)shaped over time. Again, as with extended mind on the level of sensorimotor patterns of interaction, the question of interest is not only where the boundary of the mind is located, but how the boundary expands or transforms over time. This is why the case of (particularly modern) art might be of specific interest here. The history of modern art is full of permutations of aesthetic sensibilities. One could even argue that such artworks often explicitly draw attention to and aim to reshape our aesthetic standards. The spectator is thus challenged to discover new modes of aesthetic experience and find the appropriate standard of beauty or artistic merit under which a given work of art is to be appreciated. It is through such a process that I maintain that some works of art can be said to be incorporated into our extended minds, as they come to mediate the aesthetic salience of our environment.

« 46 » The case of Otto strolling through MoMa can therefore be taken as an example that I have attempted to elucidate as a kind of externalization of bounds of viability. Of course, much could still be said about the scope and variety of this notion. I have chosen to limit my attention to a single example, mainly to clarify the two facets of EM, namely the moving boundary and plastic core, which can be thought of as quite distinct yet compatible. My example is intended to show mainly how one can be re-

garded as a "flip side" of another – how some objects may be incorporated into our operational boundary due to the pragmatic role they play in our sensorimotor interactions, while others seem to lack pragmatic value precisely because they affect the bounds of viability of those interactions – how some constitute a moving boundary, while others affect our plastic core.

Conclusion

« 47 » My findings point toward a fairly general notion of the extended mind, which to some extent unifies different frameworks involved in third-wave EM. Recall, that the main task of any such framework was to formulate a notion of a boundary of cognitive agents that exhibits plasticity and is not necessarily organism-centered.

« 48 » I have discussed two theoretical frameworks through which I addressed this challenge – predictive processing and enactivism. The two frameworks were found to be quite divergent regarding their basic points of departure – the former construes mental boundaries in terms of the Markov blanket while enactivism leads us to consider the boundaries of the mind along the lines of operational closure. With regard to this issue, I do not believe the two frameworks can be reconciled and I lean in favor of the enactivist notion of mental boundary. I claim that enactivism provides a good point of departure for third-wave EM by developing a notion of a mental boundary able to extend and undergo transformations.

« 49 » I then address the concern, that the enactivist position is necessarily organism-centered. I turn to FEP and the notions of active and perceptual inference, taken up from PP. From this vantage point, it becomes feasible to consider the notion of mind, which is plastic not only with respect to its operational boundary, but also with respect to its core structures and bounds of viability. On the issue of organism-centeredness, I claim that the frameworks of enactivism and PP may be mutually enriching. They point toward a perspective that does justice to the main commitments of third-wave EM, namely, an account of the *moving boundary* and *plastic core* of cognitive agency.

Funding

No external funding was received while writing this manuscript.

Competing interests

The author declares that he has no competing interests.

References

Allen M. & Friston K. (2018) From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. Synthese 195(6): 2459–2482. ▸ https://cepa.info/4099

Ashby W. R. (1960) Design for a brain: The origin of adaptive behaviour. Chapman and Hall, London.

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://cepa.info/4497

Chalmers D. & Clark A. (1998) The extended mind. Analysis 58(1): 10–23. ▸ https://cepa.info/4959

Clark A. (2003) Natural born cyborgs: Minds, technologies, and the future of human intelligence. Oxford University Press, Oxford.

Clark A. (2008) Supersizing the mind: Embodiment, action and cognitive extension. Oxford University Press, Oxford. Reviewed in ▸ https://constructivist.info/5/2/97

Clark A. (2016) Surfing uncertainty: Prediction, action and the extended mind. Oxford University Press, Oxford.

Clark A. (2017) How to knit your own Markov blanket: Resisting the second law with metamorphic minds. In: Metzinger T. & Wiese W. (eds.) Philosophy and predictive processing. MIND Group, Frankfurt am Main. https://predictive-mind.net/papers/papers/how-to-knit-your-own-markov-blanket

Clark A. (2018) A nice surprise? Predictive processing and the active pursuit of novelty. Phenomenology and the Cognitive Sciences 17: 521–534. https://doi.org/10.1007/s11097-017-9525-z

De Jaegher H. & Di Paolo E. A. (2007) Participatory sense-making: An enactive approach to social cognition. Phenomenology and the Cognitive Sciences 6(4): 485–507. ▸ https://cepa.info/2387

Di Paolo E. A. (2005) Autopoiesis, adaptivity, teleology, agency. Phenomenology and the cognitive sciences 4(4): 429–452.

Di Paolo E. A. (2009) Extended life. Topoi 28(1): 9–21. ▸ https://cepa.info/322

Di Paolo E. A., Cuffari E. C. & De Jaegher H. (2018) Linguistic bodies: The continuity between life and language. MIT Press, Cambridge MA. Reviewed in ▸ https://constructivist.info/15/2/194

Di Paolo E. A. & De Jaegher H. (2017) Neither individualistic nor interactionist. In: Durt C., Fuchs T. & Tewes C. (eds.) Embodiment, enaction, and culture: Investigating the constitution of the shared world. MIT Press, Cambridge MA: 87–105. ▸ https://cepa.info/2301

Di Paolo E. A., Thompson E. & Beer R. (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

Friston K. J. (2009) The free-energy principle: A rough guide to the brain? Trends in Cognitive Science 13(7): 293–301.

Friston K. J. (2013) Life as we know it. Journal of The Royal Society Interface 10(86): 20130475.

Friston K. J. & Stephan K. E. (2007) Free-energy and the brain. Synthese 159(3): 417–458.

Fuchs T. & De Jaegher H. (2009) Enactive intersubjectivity: Participatory sense-making and mutual incorporation. Phenomenology and the Cognitive Sciences 8(4): 465–486. ▸ https://cepa.info/2274

Gallagher S. (2018) The extended mind: State of the question. The Southern Journal of Philosophy 56(4): 421–447. ▸ https://cepa.info/6927

Gallagher S. & Allen M. (2018) Active inference, enactivism and the hermeneutics of social cognition. Synthese 195(6): 2627–2648. ▸ https://cepa.info/4222

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

Hohwy J. (2016) The self-evidencing brain. Nous 50(2): 259–285.

Hohwy J. (2017) How to entrain your evil demon. In: Metzinger T. & Wiese W. (eds.) Philosophy and predictive processing. MIND Group, Frankfurt am Main. https://predictive-mind.net/papers/how-to-entrain-your-evil-demon

TIMOTEJ PROSEN
has completed a master's degree in philosophy with a thesis on Husserlian phenomenology and Sellarsian philosophy of science and is currently a junior researcher and a PhD candidate at the Department of Philosophy at the University of Ljubljana, where he also works as a teaching assistant. He is working on a PhD thesis on the topic of affective experience, as conceived through the frameworks of enactivism and third-wave extended mind. His wider area of research includes process philosophy (primarily that of Whitehead, Simondon, and Deleuze), phenomenology (with a special interest in Merleau-Ponty) and non-representationalist or constructivist approaches in cognitive science (mainly enactivism).

Hohwy J. (2019) Quick'n'lean or slow and rich? Andy Clark on predictive processing and embodied cognition. In: Colombo M., Irvine E. & Stapleton M. (eds.) Andy Clark and his critics. Oxford University Press, Oxford: 191–205.

Kirchhoff M. (2012) Extended cognition and fixed properties: Steps to a third-wave version of extended cognition. Phenomenology and the Cognitive Sciences 11(3): 207–308.

Kirchhoff M. (2018) Autopoiesis, free energy, and the life–mind continuity thesis. Synthese 195: 2519–2540. ▸ https://cepa.info/4834

Kirchhoff M. & Kiverstein J. (2019) Extended consciousness and predictive processing: A third-wave view. Routledge, London.

Kirchhoff M. & Kiverstein J. (2020a) How to determine the boundaries of mind: A Markov blanket proposal. Synthese 198(8): 4791–4810.

Kirchhoff M. & Kiverstein J. (2020b) Attuning to the world: The diachronic constitution of the extended conscious mind. Frontiers in Psychology 11: 1966. https://doi.org/10.3389/fpsyg.2020.01966

Kirchhoff M., Parr T., Palacios E., Friston K. J. & 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://cepa.info/5393

Loughlin V. (2020) Extended mind. In: Glăveanu V. P. (ed.) The Palgrave encyclopedia of the possible. Springer Nature, Cham.

Menary R. (2010) Cognitive integration and the extended mind. In: Menary R. (ed.) The extended mind. MIT Press, Cambridge MA: 227–243.

Menary R. & Kirchhoff M. (2014) Cognitive transformations and extended expertise. Educational philosophy and theory 46(6): 1–14.

Merleau-Ponty M. (2005) Phenomenology of perception. Routledge, New York. French original published in 1945

O'Brien G. (1998) Review of Being There: Putting Brain, Body, and World Together Again, by Andy Clark. Metascience 7(1): 78–83.

Parr T., Pezzulo G. & Friston K. J. (2022) Active inference: The free energy principle in mind, brain, and behavior. MIT Press, Cambridge, MA.

Pearl J. (1988) Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann, Portland OR.

Ramstead M. J. D., Badcock P. & Friston K. J. (2018) Answering Schrödinger's question: A free-energy formulation. Physics of Life Reviews 24: 1–16.

Ramstead M. J. D., Kirchhoff M. D., Constant A. & Friston K. J. (2021) Multiscale integration: Beyond internalism and externalism. Synthese 198: 541–570.

Ryan K. & Schiavio A. (2019) Extended musicking, extended mind, extended agency: Notes on the third wave. New Ideas in Psychology 55: 8–17.

Seth A. K. (2015) The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In: Metzinger T. & Windt J. M. (eds.) Open MIND: 35(T). MIND Group, Frankfurt am Main. https://open-mind.net/papers/the-cybernetic-bayesian-brain

Sutton J. (2010) Exograms and interdisciplinarity: History, the extended mind, and the civilizing process. In: Richard M. (ed.) The extended mind. MIT Press, Cambridge MA: 189–225.

Szanto T. (2013) Shared extended minds: Towards a socio-integrationist account. In: Moyal-Sharrock D., Coliva A., & Munz A. M. (eds.) Mind, language and action. Proceedings of the 36th International Wittgenstein Symposium. De Gruyter, Berlin: 400–402.

Varela F. J. (1981) Autonomy and autopoiesis. In: Roth G. & Schwegler H. (eds.) Self-organizing systems: An interdisciplinary approach. Campus Verlag, Frankfurt/New York: 14–24. ▸ https://cepa.info/2063

Varela F. J. (1992) Whence perceptual meaning? A cartography of current ideas. In: Varela F. J. & Dupuy J. P. (eds.) Understanding origins: Contemporary ideas on the origin of life, mind and society. Kluwer, Boston: 235–263. ▸ https://cepa.info/2074

Varela F. J. (1997) Patterns of life: Intertwining identity and cognition. Brain and Cognition 34(1): 72–87. ▸ https://cepa.info/2010

Received: 11 February 2022
Revised: 30 May 2022
Accepted: 4 June 2022

This target article is part of a bigger picture that encompasses several open peer commentaries and the response to these commentaries. You can read these accompanying texts on the following pages (or, if you have only the target article, follow the embedded link that takes you to the journal's web page from which you can download these texts).