*Intellectica*, 2016/1, **65**, pp. 63-79

# Reconciling Constitutive and Behavioural Autonomy. The Challenge of Modelling Development in Enactive Cognition

David VERNON*

**ABSTRACT**. In the enactive paradigm of cognitive science, development plays a crucial role in the realization of cognition. This position runs counter to the computational functionalism upon which cognitivist and classical artificial intelligence systems are founded, especially the assumption that cognition can be achieved by embedding pre-formed knowledge. The enactive stance involves a progressive phased transition from cognitive capacity to cognitive capability, highlighting the role of development in extending the timescale of a cognitive agent's prospective abilities and in expanding its repertoire of effective action. We review briefly some necessary conditions for cognitive development, drawing on examples from developmental psychology, illustrating the ideas by looking at the ontogenesis of instrumental helping and collaboration in infants, and identifying some of the essential elements of a developmental cognitive architecture. We then focus on the fact that enactive systems are operationally-closed, autonomous, and self-maintaining. Consequently, there are organizational constitutive processes at play as well as behavioural ones. Reconciling these complementary processes poses a significant challenge for the creation of complete model of development that must show how constitutive autonomy is compatible with and may even give rise to behavioural autonomy. We conclude by drawing attention to recent research which could provide a way of addressing this challenge.

*Keywords*: behavioural autonomy, cognitive architecture, constitutive autonomy, develop- ment, enaction, ontogeny, phylogeny, value systems.

**RÉSUMÉ**. **Réconcilier l'autonomie constitutive et comportementale : le défi de la modélisation du processus de développement dans le paradigme de l'énaction en Sciences Cognitives**. Dans le paradigme de l'énaction en sciences cognitives, le développement joue un rôle crucial dans la réalisation de la cognition. Cette proposition va à l'encontre du fonctionnalisme computationnel et, en particulier, de l'hypothèse selon laquelle la cognition peut être obtenue en intégrant des connaissances préformées, posture sur laquelle sont fondées les approches cognitivistes et les systèmes d'intelligence artificielle classiques. La position enactive suggère une transition progressive de la capacité cognitive à la capabilité cognitive, mettant en évidence le rôle du développement dans l'extension de l'échelle de temps du pouvoir prospectif et du répertoire d'actions efficaces d'un agent cognitif. Nous passons en revue brièvement quelques conditions nécessaires pour le développement cognitif, en s'inspirant des exemples tirés de la psychologie du développement, en illustrant l'importance de l'aide instrumentale et de la collaboration dans l'ontogenèse des nourrissons, et en identifiant certains des éléments essentiels d'une architecture cognitive du développement. Nous nous concentrons sur le fait que les systèmes enactifs sont opérationnellement clos, autonomes et auto-entretenus. Par conséquent, des processus constitutifs sont en jeu tant au niveau organisationnel que

* Interaction Lab, School of Informatics, University of Skövde, Sweden. david<at>vernon.eu.

© 2016 Association pour la Recherche Cognitive.

comportemental. Concilier ces processus complémentaires pose un défi important pour la création d'un modèle complet de développement qui doit montrer comment l'autonomie constitutive est compatible avec (et peut même donner lieu à) l'autonomie comportementale. En conclusion, nous attirons l'attention sur des recherches récentes qui pourrait fournir un moyen de relever ce défi.

*Mots-clés :* Autonomie comportementale, architecture cognitive, autonomie constitutive, développement, énaction, ontogenèse, phylogénie, systèmes de valeurs.

## I – INTRODUCTION

Contemporary accounts of cognition emphasize the role of prospection in guiding the selection and execution of actions, the interdependence of action and perception, and the importance of autonomous development [1, 2]. Increasingly, too, cognition is linked to the constitutive and behavioural autonomy of a physically-embodied agent [3, 4]. These key facets of cognition are also at the core of the enactive paradigm of cognitive science [5] in which cognition subsumes action and perception rather than viewing it as distinct process that binds them together [6]. My aim in this short paper is to show how development underpins these issues and to argue that this poses a challenge for advancing our understanding of this enactive interpretation of cognition. The challenge is to understand how the value systems that drive development can support both constitutive and behavioural autonomy and how processes that support constitutive autonomy might also give rise to behavioural autonomy, as a true enactive system would require.

We begin with an overview of enaction. We then consider the issue of development from a series of perspectives, beginning with enaction itself before shifting to developmental psychology to understand the developmental process in humans and some of the motivations that drive that process. These considerations suggest the necessary elements of a cognitive architecture capable of development. Finally, we discuss how autonomy impacts on development, highlighting the challenge presented by the need to accommodate both constitutive autonomy and behavioural autonomy. We conclude by considering one possible way in with this challenge might be addressed.

## II – ENACTION

Enaction has become an increasingly-important paradigm of cognitive science with great potential for deployment in cognitive robotics [7]. Its roots can be traced to the work of Maturana in the early 1970s [8, 9], and Maturana and Varela in the late 1970s and 1980s [10, 11, 6]. It gained more widespread acceptance in the 1990s with the seminal work of Varela, Thompson, and Rosch [12], and today stands as one of the main pillars of cognitive science in the post-cognitivist era, as evidenced in the recent exposition by Stewart, Gapenne, and Di Paolo [13].

The core principle of enaction is that an autonomous cognitive system does not come pre-equipped with an established base of semantic knowledge shared by other cognitive agents but that it develops its own understanding of its world. It does this by interaction. However, the interaction does not provide "input" to determine the agent's understanding: it provides the structural

coupling of the agent with its world that conditions what is meaningful for the agent as it maintains its autonomy and enhances its ability to do so through continued development.

Enactive agents exhibit five key characteristics: autonomy, embodiment, emergence, experience, and a capacity for sense-making [14, 15]. While autonomy is a difficult concept to tie down [16] and while there are several perspectives on what it means [3], it can be viewed as the degree of self-determination of a system, *i.e.* the degree to which a system's behaviour is not determined by the environment and, thus, the degree to which a system determines its own goals [17, 18, 19, 20]. Embodiment too is a complex issue [21, 22, 23, 24]. From the perspective of enaction, the key point is that the body – its physical instantiation – plays a direct and constitutive role in cognition. It is not simply an arbitrary vehicle for symbolic information processing as it is in the cognitivist paradigm.[1] This body-dependence is central to the enactive capacity for sense-making that we discuss below. Emergence refers to the phenomenon whereby the behaviour of a system arises from the dynamic interplay between the components of the system and between the components and the system as a whole through continuous reciprocal causation (CRC) [28, 29], circular causality [10, 30], and downward causation [31, 20]. Experience is the cognitive agent's history of interaction with the world in which it is embedded and to which it is structurally-coupled. Since an enactive agent is autonomous, interactions are perturbing influences rather than inputs that control the agent: they trigger changes in the state of the agent through structural coupling, a process of mutual perturbations of the system and environment that facilitate the on-going operational identity of the agent and its autonomous self-maintenance.

This process of structural coupling produces, over time, a congruence between the system and its environment. In other words, the system develops to compensate for the threats the environmental perturbations pose to its autonomy and in the process constructs an understanding of the environment that is conditioned by its specific physical embodiment through which the structural coupling is mediated. This is what is meant by the fifth characteristic of enaction: sense-making. The knowledge, know-how, and understanding possessed by an agent is generated by the agent itself. This knowledge is not arbitrary, but captures some regularity or lawfulness in the interactions of the system, *i.e.* its experience. However, the sense the agent makes is dependent on the way in which it can interact: its own actions and its perceptions of the environment's action on it. These perceptions and actions are unique to the system itself and the resultant knowledge makes sense only insofar as it contributes to the maintenance of the system's autonomy. This view on knowledge and epistemology constrasts starkly with the computational functionalism of the cognitivist paradigm according to which there is a principled decoupling of the computational model of cognition from its instantiation as a physical system [32]. Essentially, the physical realization of a

[1] The differences between the cognitivist paradigm and the emergent paradigm (which includes enaction) are many: more than ten distinct characteristics are used to contrast them in [25] and [1]. For an overview of each see [26, 27, 2].

cognitivist computational model is inconsequential. It does not matter if it is a computer or human brain provided it supports the performance of the required symbolic computations computations.[2]

## III – ENACTION AND DEVELOPMENT

In making sense of its experience, the cognitive system is enacting – bringing forth through its actions – what is important for the continued existence of the system. This enaction is effected by the system as it is embedded in its environment, but as an autonomous entity distinct from the environment. Viewed in this light, cognition encapsulates both perception and action and is the process by which an autonomous system maintains its autonomy by leveraging its predictive capabilities. Thus, cognition can be viewed as an agent's capability to predict a need to act, the possible actions of other agents, and the outcome of all these actions. Crucially, it emerges from an innate undeveloped capacity borne of the agent's particular phylogeny and embodiment, and develops into a full-blown capability for predictive interaction. An enactive system has the capacity to increase the strength of its autonomy[3] through cognition – by making sense of its world and subsequently exploiting that acquired understanding – but the realization of the potential latent in this capacity in an operational capability requires development. In particular, development is pivotal in extending the timescale of a cognitive agent's capability for prospection and expanding its repertoire of effective action. This generative (*i.e.* self-constructed) autonomous learning and development is one of the hallmarks of the enactive approach.

In this view, development is the process of establishing and enlarging the possible space of mutually-consistent couplings in which an agent can engage without compromising its autonomy. The space of perceptual possibilities is founded not on an absolute objective environment, but on the space of possible actions that the system can engage in while still maintaining the consistency of the coupling with the environment. Through this ontogenetic development, the cognitive system develops its own epistemology, *i.e.* its own system-specific history- and context- dependent knowledge of its world, knowledge that has meaning exactly because it captures the consistency and invariance that emerges from the dynamic self-organization in the face of structural coupling with the environment. The mutual specification of the system's reality by the system and its environment through structural coupling is referred to as co-determination [6] and is related to the concept of radical constructivism [33].

This is what development entails from an enactive perspective. Let us now look at how development is achieved in practice in infants. This will provide us with some foundations for the subsequent section that sets out some of the essential elements of an enactive cognitive architecture that is capable of development and interaction with people.

---

[2] For a discussion of computational functionalism and its links to the cognitivist paradigm, see [2], Chapters 2 and 5.

[3] An explanation of distinction between strength of autonomy and degree of autonomy can be found in [2], Chapter 4.

## IV – PSYCHOLOGY AND DEVELOPMENT

The development of an autonomous agent is driven by motives and value systems [34, 35]. There are social motives and exploratory motives, reflecting, loosely, the psychology of development espoused by Vygotsky and Piaget, respectively [36, 37, 38, 39]. Both motives function from birth and provide the driving force for action throughout life.

The social motive focusses on finding comfort, security, and satisfaction through interaction with others, allowing the agent to learn new skills and acquire knowledge about the world from the experience of others. It is manifest from birth in the tendency to fixate social stimuli, imitate basic gestures, and engage in social interaction. The social motive is so important that it has been suggested that without it a person will stop developing altogether. Social motives also include a strong need to belong, a drive for self-preservation, and the need for cognitive consistency with other [40].

There are at least two exploratory motives, one to do with the discovery of novelty and regularity in the world and the other to do with finding out about the potential of one's own action capabilities. Infants are visually attracted to new objects and events but after a while they cease to be attracted. Infants also have a strong motivation to discover what they can do with objects, especially with respect to their own sensorimotor capabilities and the particular characteristics of their embodiment. Effectively, infants have a strong motivation to discover the affordances of objects around them.

The motivation to seek new ways of doing things is very strong and it can override ways of doing something that has already become established through previous development. This means that skills are developed non-monotonically: sometimes you get worse at doing something before you get better at it. So, it isn't necessarily success at achieving task-specific goals that drives development in infants but rather the discovery of new modes of interaction with the world in which the infant is embedded: the acquisition of a new way of doing something through exploration [41, 42].

In developing new skills and in learning how to act and interact, prospection comes to the fore. Actions are initiated and executed by a motivated subject and they are defined by the goal of the action, not the specific movements by which the goal is achieved. More especially, they are guided by prospective information [1]. For example, when performing manipulation tasks or observing someone else performing them, people fixate on the goals and sub-goals of the movements, *e.g.* the point where an object is to be grasped, the target location of object, and the support surface, not on the body parts, *e.g.* the hands or the grasped object. In other words, gaze is governed by predictive motor control. Again, we see that development is focussed on expanding the repertoire of actions and extending the time horizon of an agent's predictive capacity.

The phased aspect of development is particularly relevant in the manner in which infants and children come to understand the intentions of others and to help them achieve their goals. It takes several years for human infants to develop the requisite abilities.

During the first year of life the progressive acquisition of motor skills facilitates the development of an ability to understand the intentions of other agents, initially by anticipating the goal of simple movements and eventually understanding more complex goals. During this period, the ability to infer what another agent is focussing their attention on and the ability to interpret emotional expressions begins to improve substantially. Around 14 to 18 months of age children begin to exhibit instrumental helping behaviour, *i.e.* they display spontaneous, unrewarded helping behaviours when another person is unable to achieve his goal [43]. This is a critical stage in the development of a capacity for collaborative behaviour, a process that progresses past three and four years of age. Around 2 years of age children start to solve simple cooperation tasks together with adults [44]. This phase of development sees the beginning of shared intentionality where a child and an adult form a shared goal and both engage in joint activity. Children seem to be motivated not just by the goal but by the cooperation itself, *i.e.* the social aspect of the interaction. The ability to cooperate with peers and become a social partner in joint activities develops over the second and third years of life as social understanding increases [45]. More complex collaboration, which necessitates the sharing of intentions and joint coordination of actions, appears at about three years of age when children master more difficult cooperation tasks such as those involving complementary roles for the two partners in a collaborative task [46]. At three years of age, children begin to develop the ability to cooperate by coordinating two complementary actions. By three-and-a-half years of age children quickly master the task, can deal effectively with the roles in the task being reversed, and can even teach new partners [47]. The motives which drive instrumental helping are simpler than those of collaborative behaviours: they are based on wanting to see the goal completed or wanting to perceive pleasure in the human at being able to complete it. In this case, the motivational focus is solely on the needs of the second agent and the needs of the first agent do not figure in this. The motives underlying collaborative behaviour are more complicated. In this case, the intentions and the goals have to be shared and the motivational focus is on the needs of both agents.

The foregoing has provided some insights into the ontogenetic aspect of development, *i.e.* the developmental process itself and, consequently, it suggests some of the social and exploratory elements that an agent must possess for development to happen. Drawing on this, let us now look at development from the perspective of phylogeny, *i.e.* the agent's cognitive architecture.

## V – COGNITIVE ARCHITECTURES AND DEVELOPMENTS

While the term cognitive architecture derives from Allen Newell's pioneering work in cognitivist cognitive science, and in particular to his work and his colleagues work on unified theories of cognition [48, 49, 50], it is also used by those who work in enactive systems to refer to the phylogenetic configuration of a new-born or newly-created cognitive agent: the initial state from which it subsequently develops. An appropriately-configured cognitive architecture doesn't guarantee successful development because, as we saw in the previous section, development also requires exposure to an environment

that is conducive to development, one in which there is sufficient regularity to allow the system to build a sense of understanding of the world around it, but not excessive variety that would overwhelm an agent which has inherent limitations on the speed with which it can develop. Thus, cognition has two necessary elements: phylogeny and ontogeny, *i.e*. a cognitive architecture and gradually-acquired experience.

Although several guidelines for configuring cognitive architectures have been proposed, *e.g.* [51, 52, 53, 54, 55], few address development explicitly, mainly because these guidelines derive from work in cognitivist cognitive science. On the other hand, Jeffrey Krichmar proposes five design principles for developmental artificial brain-based devices [56, 57, 58] which are also applicable to cognitive architectures.

- First, the cognitive architecture should address the dynamics of the neural elements in different regions of the brain, the structure of these regions, and especially the connectivity and interaction between these regions. In other words, a developmental cognitive architecture should make explicit the operation of the system as a whole.
- Second, the architecture should support perceptual categorization: *i.e.* the capacity to organize unlabelled sensory signals of all modalities into categories without prior knowledge or external instruction. In effect, this means that the system should be autonomous and, as a developmental system, it should be a model generator, rather than a model fitter (a point also emphasized by John Weng [59]).
- Third, a developmental system must have a physical instantiation, *i.e.* it must be embodied, with the system's morphology conditioning the agent's understanding of its environment.
- Fourth, the system should have some minimal set of innate behaviours or reflexes in orderto explore and survive in its initial environmental niche. From this minimum set, the system can develop so that it improves its behaviour over time.
- Fifth, and of particular importance to the argument in this article, a developmental system should have a means to adapt. This entails the presence of a value system, *i.e.* a set of motivations that guide or govern its development [34, 35]. These should be non-specific (in the sense that they don't specify what actions to take) modulatory signals that bias the dynamics of the system so that the global needs of the system are satisfied: in effect, so that the system's autonomy is preserved or enhanced.

Directly or indirectly, these value systems should manifest the social motives that enable fixation on social stimuli, imitation of basic gestures, and engagement in social interaction, and exploratory motives that facilitate the discovery of novelty and regularities in the environment and the system's own action capabilities, in line with the brief synopsis of infant development in the previous section.

## VI – AUTONOMY AND DEVELOPMENT

So far, so good. However, enactive cognitive systems are, first and foremost, autonomous systems. As noted already, autonomy is a difficult concept to tie down [16] and there are several perspectives on what it means [3]. Nonetheless, few would disagree that autonomy is linked to degree of self-determination of a system, *i.e*. the degree to which a system's behaviour is not determined by the environment and, thus, the degree to which a system determines its own goals [17, 18, 19, 20]. For biological autonomous agents, as well as bio-inspired artificial agents, the issue of autonomy is one of survival in the face of precarious conditions, operating in an uncertain possibly-dangerous constantly-changing environment. To do this, it must keep itself intact as an autonomous system, both physically and organizationally as a dynamic self-sustaining entity. The self-maintenance of autonomy *is* a crucial aspect of enactive cognitive agents [60], continually repairing damage to itself. Since it is better if the agent can avoid damage in the first place, cognition, as a prospective modulator of perception and action, is one of the primary mechanisms at the agent's disposal [4] to anticipate the need for action and the outcome of that action.

From this perspective, autonomy, aided by cognition, is the self-maintaining organizational characteristic of living creatures that enables them to use their own capacities to manage their interactions with the world in order to remain viable [61]. In other words, autonomy is the process by which a system manages – self-regulates – to maintain itself as a viable entity despite the precarious conditions with which the environment continually confronts it. Arguably, autonomy and autonomy-preserving processes are the foundation of cognition [60].

While more than twenty types of autonomy can be distinguished [2], two broad classes can be discerned: behavioural autonomy and constitutive autonomy [3, 4]. Behavioural autonomy is concerned with the external behaviour of the system: the extent to which the agent sets its own goals and its robustness and flexibility in achieving them as it interacts with the world around it, including other cognitive agents. Constitutive autonomy is concerned with the internal organization and the organizational processes that keep the system viable, maintaining itself as an identifiable autonomous entity. Indeed, Maturana and Varela, whose work provided the inspiration for the enactive view of cognition, define autonomy as "the condition of subordinating all changes to the maintenance of the organization" [11]. Constitutive autonomy and behavioural autonomy are related: an agent can not deal with uncertainty and danger if it is not organizationally – constitutively – equipped to do so. Behaviour depends on internal preparedness but appropriate behavioural is needed to allow the agent to achieve the requisite environmental conditions – through interaction – for constitutive autonomy to be able to operate effectively. This complementarity of the constitutive and the behavioural reflects two different sides of the characteristic of recursive self-maintenant systems [60] to deploy different processes of self-maintenance depending on environmental conditions, with constitutive and behavioural autonomy corresponding to the internal – endogenous – and external – exogenous – aspects of that adaptive capacity, respectively.

Self-regulation is central to constitutive autonomy. In biological systems, the automatic regulation of physiological functions is referred to as *homeostasis* [62, 63]: "the process of maintaining the internal milieu physiological parameters (such as temperature, pH and nutrient levels) of a biological system within the range that facilitates survival and optimal function" [64, 65]. It has been suggested [66, 67] that the autonomy of an agent is effected through a hierarchy of homeostatic self-regulatory processes, exploiting a progression of associated affective (*i.e.* emotional or feeling) states, ranging from basic reflexes linked to metabolic regulation, through drives and motives, and on to the emotions and feelings often linked to higher cognitive functions, similar to Damasio's hierarchy of levels of homeostatic regulation [64].

Typically, the autonomous agent is perturbed during interactions with the world with the result that the organizational dynamics have to be adjusted. This process of adjustment is exactly what is meant by homeostasis and the motives at every level of this hierarchy of homeostatic processes are effectively the drives that are required to return the agent to a state where its autonomy is no longer threatened. In the interaction with the world around it, the perturbations of the agent by the environment have no intrinsic value in their own right: they are just the stuff that happens to the agent as it goes about its business of survival. However, for the agent this stuff – these interactions and perturbations – has a perceived value in that it acts to endanger or support its autonomy. This value is conveyed through the affective aspect of these homeostatic processes and consequently the agent then attaches some value to what is an otherwise neutral world (even if it is a precarious one) [68]. This gives rise to a reciprocal coupling – and mutual dependency – of action and perception in cognition where perceptions and actions form a complementary set of environment-agent / agent-environment perturbations that are related not as extrinsic stimulus-response perceptuo-motor contingencies but as intrinsic processes that lead to the regulation of the system and autonomy preservation through emergent self-organization [69]. The processes of perception and action are mutually dependent because they are both modulated by the system – globally-determined – through downward causation [31, 20] and, together with other homeostatic processes, they give rise to the global constitutive autonomy-preserving system behaviour. This is a subtle but important point as it suggests a causal link between the processes of constitutive autonomy (*qua* self-organization) and behavioural autonomy (*qua* viable interaction with the environment). We return to this point in the next section.

Just as, from the perspective of behavioural autonomy, a cognitive agent continually deploys prospection through internal simulation to prepare to act [70, 71, 72, 73, 74], so too are the processes of constitutive autonomy prospective. This predictive self-regulation is known as *allostasis* [75, 76, 77]. Sterling notes that allostasis provides a *global* mechanism for overriding normal homeostasis, serving the organism as a whole with the resources previously learned to be necessary to meet predicted environmental pressures [75]. Thus, allostasis differs from homeostasis in its predictive character and in its ability to anticipate and adapt to change rather than resist it. Significantly,

allostasis is effected at a higher level of organization, involving greater number of sub-systems acting together in a coordinated manner with global processes modulating local ones, reflecting the character of circular causality. In contrast, mechanisms for homeostasis operate at a simpler level of negative feedback control [75, 77, 78].

Now here we finally come to the challenge. Development is commonly cast as a process of adaptation based on interaction with the environment and other agents [7, 1] and, when autonomy is considered, the focus is usually on behavioural autonomy. However, here we see the critical importance of constitutive autonomy. Development applies not only to the behavioural capacities for interaction that we have discussed above, but it applies also to the constitutive elements of the agent. Specifically, the value systems that drive development are relevant not just to the processes of behavioural autonomy but also to those of constitutive autonomy and both forms of autonomy exhibit prospection, the key attribute of cognition. We have seen how prospection is essential for effective processes of behaviour and interaction (*e.g.* we anticipate the need to buy groceries before cooking dinner) but it is also essential for effective constitutive processes (*e.g.* blood sugar levels are raised in anticipation of the demands of imminent exercise; see [75, 77] for this and other examples of predictive metabolic regulation). Furthermore, since enactive systems are operationally-closed,[4] autonomous, and self-maintaining, the constitutive processes may be the primary source of autonomy for both constitutive processes and behavioural interaction processes. Since development is normally cast in behavioural terms, *i.e.*, in terms of an agent interacting with the world around it, not in terms of internal interaction, this presents us with a dilemma: how can the value systems and motivations that drive development support both constitutive autonomy and behavioural autonomy? In particular, how can the processes that support constitutive autonomy also give rise to behavioural autonomy and especially to the development of the agent based on interaction with its environment and other agents through social and exploratory motives?

## VII – ADRESSING THE CHALLENGE

While the aim of this article is to identify the challenge of modelling development in enactive cognitive agents, in general, and enactive cognitive robots, in particular, it would be rather unsatisfactory to finish without suggesting where possible answers might be sought. We do this now.

[4] The term operational closure characterizes any system that is identified by an observer to be self-contained and parametrically-coupled with its environment but not controlled by the environment. It is related to organizational closure, a necessary characteristic of a particular form of self-producing self-organization called autopoiesis [8, 9]. Technically, autopoiesis operates at the bio-chemical level, *e.g.*, in cellular systems, but its usage has been expanded to deal with autonomous systems in general where, more correctly, it is referred to as operational closure. The operational closure vs. organizational closure terminology can be confusing because in some earlier publications, *e.g.* [10], Varela refers to organizational closure but in later works (by Maturana and Varela themselves, *e.g.* [6], and by others, *e.g.* [13]) this term was subsequently replaced in favour of operational closure to reflect is more general usage, with organizational closure being used to characterized an operationally-closed system that exhibits some form of self-production or self-construction [79].

Any substantive answers to the questions raised above will probably be founded on an innate capacity for allostatic and homeostatic self-organization that makes sense of the agent's structural coupling with its environment in maintaining the agent's constitutive and behavioural autonomy. It will trade the traditional emphasis on exteroception for interoception and internal action, much as John Weng has suggested with his self-affecting self-effecting models of autonomous mental development [59], and it will leverage the innate phylogenetic capacities for modulating its interactions as set out in Section 4 and 5 above.

In recent work [80, 81], Anil Seth discusses the importance of prediction in cognition, suggesting, as others have done [82], that the brain engages in continual predictive inference of the causes of sensory perturbations, *i.e.* the predictive perception of sensorimotor contingencies. In this, he develops the concept of *predictive processing* whereby the brain infers the most likely causes of its sensory inputs by minimizing the difference between sensory signals and signals derived from continuously updated predictive models. However, his central thesis is that this process derives less from classical exteroception than from interoception. This interception is based on cybernetic principles. They assert that

> "the purpose of cognition (including perception and action) is to maintain the homeostasis of essential variables and of internal organization... [and]... perception emerges as a *consequence* of a more fundamental imperative towards organizational homeostasis, and not as a stage in some process of internal world-model construction" [81], p. 8

Viewed in this enactive light, cognitive agents adapt – develop – to ensure continued existence by successfully responding to environmental perturbations so as to maintain their internal organization. This neatly links constitutive autonomy to behavioural autonomy and suggests how behavioural autonomy derives from constitutive autonomy. The question then is: how is this accomplished?

Seth builds on Karl Friston's *Free Energy Principle* [83, 84], according to which "organisms obey a fundamental imperative towards the avoidance of (information-theoretically) surprising events, according to which they must minimize in the long-run average surprise of sensory states, since surprising sensory states are (in the long run) likely to reflect conditions incompatible with continued existence" [81], p. 2. Seth suggests that *active inference*, an extension of predictive processing, operates to suppress the interoceptive prediction errors not only by updating the generative model that gave rise to the predictions but by internal action, translating the predictions into reference points for autonomic regulatory processes, *e.g.* physiological organizational homeostasis.[5] He notes that attention can then be viewed as a way of balancing

[5] Seth views allostasis as "the process of achieving homeostasis" [81], p.7, emphasizing its roots in cybernetics, in general, and the ultrastability of Ashby's homeostat [85, 86, 87, 88], in particular. He notes that the fundamental cybernetic principle is for systems to ensure their continued existence by successfully responding to environmental perturbations so as to maintain their internal organization. He goes further, stating that "The purpose of cognition (including perception and

active inference and model update, referred to as precision weighting. He reinforces the idea that "an organism should maintain well-adapted predictive models of its own physical body [...] and of its internal physiological condition" [80], p. 567. Active inference can act both to selectively sample sensory data to conform to current predictions and to seek evidence that contradicts current predictions or disambiguate multiple competing hypotheses. This leverages "the capacity of predictive models to encode counterfactual relations linking potential (but not necessarily executed) actions to their expected consequences". It implies model comparison and selection, not just the optimization of the parameters of a single model, much as the HAMMER architecture for internal simulation in cognitive robotics does with its multiple forward and inverse models [89, 90].

It remains, however, to find a way of minimizing the information-theoretic surprisal associated with the agent's internal organization. One possibility is an information-theoretic technique introduced by Robert Ulanowicz [91, 92, 93]. Although the model was originally intended to model the growth and development of ecosystems, it is quite general and has already been used to characterize the emergence and development of beliefs in human cognition [94] and is also being investigated a value system to drive the development of joint episodic-procedural memory networks [95]. Modelling the system as a flow network, growth and development are framed as a variational principle asserting that self-sustaining – autonomous – systems tend to self-organize so as to optimize what Ulanowicz refers to as *internal ascendancy*, a function of the average mutual information in the network. Systems that self-organize in this way increase the order of the network and reduce the average surprisal (an information-theoretic measure) in the network. Crucially, Ulanowicz points out that this ordering process must be balanced against the need to retain some residual entropy in the system so that it has a capacity to adapt to unpredictable perturbations. Hence, he posits the principal of *optimal* ascendancy, rather than *maximum* ascendancy: enough order and organization (captured by his information-theoretic measure of ascendancy) to facilitate prediction but enough disorder (entropy) to allow alternative to be self-selected if there is a failure in some part of the system. The balance between the two is dynamic and depends on the long-term variability of the environmental perturbations. Environments that are less precarious and less prone to change will result in a balance biased towards greated ascendancy since it is well matched by an internal organization with low surprisal; *vice versa*, an environment that is highly uncertain needs to be matched by a system that has a capability to deal with a greater degree of surprisal and, hence, has more entropy in its organization. This self-organizing tendency towards optimal ascendancy – average mutual information in network flow – parallels Friston's free energy principle and the avoidance of surprise. Together, along with Seth's interoceptive adaptive inference, they present a plausible way forward in resolving the dilemma of development in enactive cognition, one that, by focussing on processes of constitutive autonomy and by linking them to consequent processes of behavioural autonomy, is faithful to the roots of

action) is to maintain the homeostasis of essential variables and of internal organization (ultrastability)" [81], p. 8.

enaction and its foundations in operationally-closed self-sustaining autonomous systems that are structurally-coupled to their ecological niche in the environment and co-determined by it.

**ACKNOWLEDGEMENTS**

The overview of the development of an infant's ability to engage in instrumental helping and to collaborate with others follows closely the treatment in [2], p. 140, based on material compiled by Alessandra Sciutti, Istituto Italiano di Tecnologia (IIT) and the work of Claes von Hofsten, Uppsala University, Sweden, and Harold Bekkering, Radboud University Nijmegen, The Netherlands.

**REFERENCES**

[1] Vernon, D., von Hofsten, C. & Fadiga, L. (2010). *A Roadmap for Cognitive Development in Humanoid Robots*, Cognitive Systems Monographs (COSMOS), vol. 11. Berlin: Springer.

[2] Vernon, D. (2014). *Artificial Cognitive Systems – A Primer*. Cambridge, Mass.: The MIT Press.

[3] Froese, T., Virgo, N. & Izquierdo, E. (2007). Autonomy: a review and a reappraisal. In F. Almeida e Costa, L. Rocha, E. Costa, I. Harvey & A. Coutinho (eds.). *Proceedings of the 9th European Conference on Artificial Life: Advances in Artificial Life*, volume 4648, pp. 455-465, Heidelberg: Springer.

[4] Barandiaran, X. & Moreno, A. (2008). Adaptivity: From metabolism to behavior. *Adaptive Behavior*, *16(5)*, 325-344.

[5] Barandiaran, X.E. (2016). Autonomy and enactivism: Towards a theory of sensorimotor autonomous agency. *Topoi*, 1-22, March 2016.

[6] Maturana, H. & Varela, F. (1987). *The Tree of Knowledge – The Biological Roots of Human Understanding*. Boston & London: New Science Library,

[7] Vernon, D. (2010). Enaction as a conceptual framework for development in cognitive robotics. *Paladyn Journal of Behavioral Robotics, 1(2)*, 89_98.

[8] Maturana, H. (1970). *Biology of cognition*. Research Report BCL 9.0, Urbana, Illinois: University of Illinois.

[9] Maturana, H. (1975). The organization of the living: a theory of the living organization. *Int. Journal of Man-Machine Studies*, *7(3)*, 313-332.

[10] Varela, F. (1979). *Principles of Biological Autonomy*. New York: Elsevier North Holland.

[11] Maturana; H.R. & Varela, F.J. (1980). *Autopoiesis and Cognition – The Realization of the Living. Boston Studies on the Philosophy of Science*. Dordrecht, Holland: D. Reidel Publishing Company,.

[12] Varela, F., Thompson, E. & Rosch, E. (1991). *The Embodied Mind.* Cambridge, Mass.: The MIT Press.

[13] Stewart, J., Gapenne, O. & Di Paolo, E.A. (2010). *Enaction: Toward a New Paradigm for Cognitive Science*. Cambridge, Mass.: The MIT Press.

[14] Di Paolo, E., Rohde, M. & De Jaegher, H. (2010). Horizons for the enactive mind: Values, social interaction, and play. In J. Stewart, O. Gapenne, & E. Di Paolo (ed.) *Enaction: Towards a New Paradigm for Cognitive Science*, (pp. 33-87). Cambridge, Mass.: The MIT Press.

[15] Thompson E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Boston: Harvard University Press.

[16] Boden, M.A. (2008). Autonomy: What is it? *BioSystems*, *91*, 305-308.

[17] Ziemke, T. (1997. The 'environmental puppeteer' revisited: A connectionist perspective on 'autonomy'. In *Proceedings of the 6th European Workshop on Learning Robots,* Brighton, UK, August 1997.
[18] Ziemke, T. (1998). Adaptive behaviour in autonomous agents. *Presence*, *7(6)*, 564-587.
[19] Bertschinger, N., Olbrich, E., Ay, N. & Jost, J. (2008). Autonomy: An information theoretic perspective. *Biosystems*, *91(2)*, 331-345.
[20] Seth, A. (2010). Measuring autonomy and emergence via Granger causality. *Artificial Life*, *16(2)*, 179-196.
[21] Chrisley, R. & Ziemke, T. (2002). Embodiment. In *Encyclopedia of Cognitive Science* (pp. 1102-1108). London: Macmillan.
[22] Anderson, M.L. (2003). Embodied cognition: A field guide. Artificial I*ntelligence*, *149(1)*, 91-130.
[23] Anderson, M. (2007). How to study the mind: An introduction to embodied cognition. In F. Santoianni & C. Sabatano (eds.) *Brain Development in Learning Environments: Embodied and Perceptual Advancements* (pp. 65-82). Cambridge Scholars Press.
[24] Shapiro, L. (2011). *Embodied Cognition*. London: Routledge.
[25] Vernon, D., Metta G. & Sandini, G. (2007). A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation*, *11(2)*, 151-180.
[26] Varela, F.J. (1992). Whence perceptual meaning? A cartography of current ideas. In F.J. Varela & J.-P. Dupuy (eds.) *Understanding Origins – Contemporary Views on the Origin of Life, Mind and Society, Boston Studies in the Philosophy of Science* (pp. 235-263) Dordrecht: Kluwer Academic Publishers.
[27] Clark, A.(2001). *Mindware – An Introduction to the Philosophy of Cognitive Science*. New York: Oxford University Press.
[28] Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, Mass.: The MIT Press.
[29] Clark, A (1998). Time and mind. *Journal of Philosophy*, *XCV(7)*, 354-376.
[30] Kelso, J.A.S. (1995). *Dynamic Patterns – The Self-Organization of Brain and Behavior*. Cambridge, Mass.: The MIT Press, 3rd edition.
[31] Thompson, E. & Varela, F. (2001). Radical embodiment: neuronal dynamics and consciousness. *Trends in Cognitive Sciences*, *5*, 418-425.
[32] Piccinini, G. (2010). The mind as neural software? Understanding functionalism, computationalism, and computational functionalism. *Philosophy and Phenomenological Research*, *81(2)*, 269-311, September 2010.
[33] Glaserfeld, E. von (1995). *Radical Constructivism*. London: Routledge Falmer.
[34] Merrick, K E. (2010). A comparative study of value systems for self-motivated exploration and learning by robots. *IEEE Transactions on Autonomous Mental Development*, *2(2)*, 119-131, June 2010.
[35] Oudeyer, P.-Y., Kaplan, F. & Hafner, V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, *11(2)*, 265-286.
[36] Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic Books.
[37] Vygotskyn, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Mass.: The MIT Press.
[38] Lindblom, J. & Ziemke, T. (2003). Social situatedness of natural and artificial intelligence: Vygotsky and beyond. *Adaptive Behavior*, *11(2)*, 79-96.
[39] Lindblom, J. (2015). *Embodied Social Cognition,* volume 26 of *Cognitive Systems Monographs (COSMOS*). Berlin: Springer,

[40] Forgas, J.-P., Kipling, K.D. & Laham, S.M. (eds.) (2005). *Social Motivation*. Cambridge University Press.
[41] Hofsten, C. von. (2003) On the development of perception and action. In J. Valsiner & K.J. Connolly (eds.) *Handbook of Developmental Psychology* (pp. 114-140). London: Sage.
[42] Hofsten, C. von (2004). An action perspective on motor development. *Trends in Cognitive Sciences*, *8*, 266-272.
[43] Warneken, F. & Tomasello, M. (2009). The roots of human altruism. *British Journal of Psychology*, *100(3)*, 455-471.
[44] Warneken, F., Chen, F. & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development*, *77*, 640-663.
[45] Brownell, C.A., Ramani, G.B. & Zerwas, S. (2006). Becoming a social partner with peers: cooperation and social understanding in one- and two-year-olds. *Child Development*, *77(4)*, 803-821.
[46] Meyer, M., Bekkering, H., Paulus, M. & Hunnius, S (2010). Joint action coordination in 2½- and 3-year-old children. *Frontiers in Human Neuroscience*, *4(220)*, 1-7.
[47] Ashley, J. & Tomasello, M. (1998). Cooperative problem solving and teaching in preschoolers. *Social Development*, *7*, 143-163.
[48] Newell, A. (1982). The knowledge level. *Artificial Intelligence*, *18(1)*, 87-127, March.
[49] Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, Mass.: Harvard University Press,
[50] Anderson, J.R., Bothell, D., Byrne M.D., Douglass, S., Lebiere, C. & Qin Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 1036-1060.
[51] Sloman, A. (2001). Varieties of affect and the cogaff architecture schema. In *Proceedings of the AISB '01 Symposium on Emotion, Cognition, and Affective Computing*, York, UK.
[52] Sun, R. (2004). Desiderata for cognitive architectures. *Philosophical Psychology*, *17(3)*, 341-373.
[53] Hawes, N., Wyatt, J. & Sloman, A. (2006). An architecture schema for embodied cognitive systems. In *Technical Report CSR-06-12*. University of Birmingham, School of Computer Science.
[54] Langley, P., Laird, J.E. & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, *10(2)*, 141-160.
[55] Sun, R. (2007). The importance of cognitive architectures: an analysis based on clarion. *Journal of Experimental & Theoretical Artificial Intelligence*, *19(2)*, 159-193.
[56] Krichmar, J.L. & Edelman, G.M. (2005). Brain-based devices for the study of nervous systems and the development of intelligent machines. *Artificial Life*, *11*, 63-77.
[57] Krichmar, J.L. & Reeke, G.N. (2005). The Darwin brain-based automata: Synthetic neural models and real-world devices. In G.N. Reeke, R.R. Poznanski, K.A. Lindsay, J.R. Rosenberg, & O. Sporns (eds.) *Modelling in the neurosciences: from biological systems to neuromimetic robotics* (pp. 613-638). Boca Raton, Taylor and Francis.
[58] Krichmar, J.L. & Edelman, G.M. (2006). Principles underlying the construction of brain-based devices. In T. Kovacs & J.A.R. Marshall (eds.), *Proceedings of AISB '06 - Adaptation in Artificial and Biological Systems*, volume 2 of *Symposium on Grand Challenge 5: Architecture of Brain and Mind* (pp. 37-42). Bristol: University of Bristol.
[59] Weng, J. (2004). Developmental robotics: Theory and experiments. *International Journal of Humanoid Robotics*, *1(2)*, 199-236.

[60] Bickhard, M.H. (2000). Autonomy, function, and representation. *Communication and Control-Artificial Intelligence*, *17(3-4)*, 11-131.
[61] Christensen, W.D. & Hooker, C.A. (2000). An interactivist-constructivist approach to intelligence: self-directed anticipative learning. *Philosophical Psychology*, *13(1)*, 5-45.
[62] Cannon, W.B. (1929). Organization of physiological homeostasis. *Physiological Reviews*, *9*, 399-431.
[63] Bernard, C (1878). *Leçons sur les phénomènes de la vie communs aux animaux et végétaux*. Paris: J.-B. Baillière.
[64] Damasio, A.R. (2003). *Looking for Spinoza: Joy, sorrow and the feeling brain*. Orlando, Florida: Harcourt.
[65] Damasio, A. & Carvalho, G.B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews Neuroscience*, *14*, 143-152.
[66] Morse, A., Lowe, R. & Ziemke, T. (2008). Towards an enactive cognitive architecture. In *Proceedings of the First International Conference on Cognitive Systems*, Karlsruhe:, Germany.
[67] Ziemke, T. & Lowe, R. (2009). On the role of emotion in embodied cognitive architectures: From organisms to robots. *Cognition and Computation*, *1*, 104-117.
[68] Di Paolo, E.A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4(4)*, 429-452.
[69] Vernon, D., Lowe, R., Thill, S. & Ziemke, T. (2015). Embodied cognition and circular causality: On the role of constitutive autonomy in the reciprocal coupling of perception and action. *Frontiers in Psychology*, *6(1660)*, 1-13, October 2015.
[70] Schacter, D.L., Addis, D.R. & Buckner, R.L. (2008). Episodic simulation of future events: Concepts, data, and applications. *Annals of the New York Academy of Sciences*, *1124*, 39-60.
[71] Seligman, M.E.P., Railton, P., Baumeister, R.F. & Sripada C.(2013). Navigating into the future or driven by the past. *Perspectives on Psychological Science*, *8(2)*, 119-141.
[72] Hesslow, G.(2012). The current status of the simulation theory of cognition. *Brain Research*, *1428*, 71-79.
[73] Svensson, H., Lindblom, J. & Ziemke, T. (2007). Making sense of embodied cognition: Simulation theories of shared neural mechanisms for sensorimotor and cognitive processes. In T. Ziemke, J. Zlatev & R.-M. Frank (eds.) *Body, Language and Mind*, volume 1: *Embodiment* (pp. 241-269). Berlin: Mouton de Gruyter,
[74] Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*, 377-442.
[75] Sterling, P. (2004). Principles of allostasis. In J. Schulkin (ed.) *Allostasis, Homeostasis, and the Costs of Adaptation* (pp. 17-64). Cambridge University Press.
[76] J. Schulkin (2011). Social allostasis: anticipatory regulation of the internal milieu. *Frontiers in evolutionary neuroscience*, *2(111)*, 1-15.
[77] Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology and Behaviour*, *106(1)*, 5-15.
[78] Muntean, I. & Wright, C.D. (2007). Autonomous agency, AI, and allostasis – a biomimetic perspective. *Pragmatics & Cognition*, *15(3)*, 485-513.
[79] Froese, T. & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, *173*, 466-500.
[80] Seth, A.K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17(11)*, 565–573, November 2013.

[81] Seth, A.-K. (2015). The cybernetic Bayesian brain – from interoceptive inference to sensorimotor contingencies. In T. Metzinger & J.-M. Windt (eds.) *Open MIND*, volume 35, (pp. 1-24). Frankfurt am Main: MIND Group.
[82] Downing, K. (2009). Predictive models in the brain. *Connection Science*, *21*, 39-74.
[83] Friston, K.-J. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13(7)*, 293-301.
[84] Friston, K.J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11(2)*, 127-138.
[85] Ashby, W.R. (1952). *Design for a Brain*. New York: John Wiley & Sons, first edition.
[86] Ashby, W.R. (1954). *Design for a Brain*. New York: John Wiley & Sons, first edition. Reprinted with corrections.
[87] Ashby, W.R. (1960). *Design for a Brain*. New York: John Wiley & Sons, second edition.
[88] Vernon, D. (2013). Interpreting Ashby – but which one? *Constructivist Foundations*, *9(1)*, 111-113, November 2013.
[89] Demiris, Y. & Hayes, G. (2002). Imitation as a dual-route process featuring predictive learning components: a biologically-plausible computational model. In K. Dautenhahn & C. Nehaniv (eds.) *Imitation in Animals and Artifacts*, Chapter 13 (pp. 327-361). Cambridge, Mass.: The MIT Press.
[90] Demiris, Y. & Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition (HAMMER*). Robotics and Autonomous Systems*, *54*, 361-369.
[91] Ulanowicz, R.E. (1998). A phenomenology of evolving networks. *Systems Research and Behavioural Science*, *15*, 373-383.
[92] Ulanowicz, R.E. (2000). *Growth and Development; Ecosystems Phenomenology*. Lincoln, Nebraska: toExcel Press,
[93] Ulanowicz, R.E. (2011). Quantitative methods for ecological network analysis and its application to coastal ecosystems. In E. Wolanski & D. S. McLusky (eds.) *Treatise on Estuarine and Coastal Science*, volume 9 (pp. 37-57). Waltham, Mass.: Academic Press.
[94] Castillo, R.D., Kloos, H., Richardson, M.J. & Waltzer, T. (2015). Beliefs as self-sustaining networks: Drawing parallels between networks of ecosystems and adults' predictions. *Frontiers in Psychology*, *6*, 1723.
[95] Vernon, D., Beetz, M. & Sandini, G. (2015). Prospection in cognitive robotics: The case for joint episodic-procedural memory. *Frontiers in Robotics and AI*, *2* (Article 19), 1-14.