The only wrong cell is the dead one: On the enactive approach to normativity

Manuel Heras-Escribano[1], Jason Noble[2] and Manuel de Pinedo[1]

[1] Departamento de Filosofía I, Universidad de Granada, Spain
[2] School of Electronics and Computer Sciences, University of Southampton, UK
herasescribano@gmail.com

Abstract

In this paper we challenge the notion of 'normativity' used by some enactive approaches to cognition. We define some varieties of enactivism and their assumptions and make explicit the reasoning behind the co-emergence of individuality and normativity. Then we argue that appealing to dispositions for explaining some living processes can be more illuminating than claiming that all such processes are normative. For this purpose, we will present some considerations, inspired by Wittgenstein, regarding norm-establishing and norm-following and show that attributions of normativity to non-social agents are deeply paradoxical. The main conclusions of our discussion are: (1) circular and internal explanations centred on the stability of living systems are insufficient to account for processes where the environment plays an important role, such as adaptation. Enactivism is not an explanatory alternative to evolutionary biology but needs it as a complement to accounts focused on the internal self-assembly of organisms; (2) though we share enactivism's anti-representational spirit, we argue that ecological psychology can offer a better account of perception.

Enactivism's natural norms

Enactivism is often presented as the new paradigm for explaining cognition (Stewart et al., 2010). It is based on the assumption that cognition, rather than being a matter of abstract calculus and manipulation of internal representations in the head, is a spatio-temporally extended and dynamical process in which an embodied agent is meaningfully dealing with its environment in order to adapt itself to it. But this is not enough for defining enactivism. In fact, all these assumptions and tools have already been endorsed and developed by other anti-cognitivist theories, such as ecological psychology (Gibson, 1966, 1979). So, what's new about enactivism? For some enactivists (Di Paolo, 2009) their theory provides a definition of agency in which the new embodied, extended and anti-representational cognitive science can rely on. Enactivism was born as a biological theory that emphasized the continuity between life and mind (Canguilhem, 1965; Maturana and Varela, 1987). Among the impressive achievements of the theory, perhaps the most important was the change of focus in thinking about living creatures: these are not seen as mere compounds of parts selected by evolution, but as whole agents individuated from their environment in terms of their internal structure. This structure or system is based on different networked processes (such as metabolism and the different processes of the nervous system, for example) and it is taken as a unity. This is to say that the system as a whole provides stability, and the processes of this system that result from its stable configuration are intended to keep this unity going. An agent, thus, is autonomous or self-sustained, and its goal is to keep this self-stability. This is the sense in which life is normative, according to enactivism. Although this idea was already being embraced by some philosophers (Canguilhem, 1965; Jonas, 1966) only the explicit analysis due to Maturana and Varela made it into a suitable starting point for thinking about cognition. Cognition is one species in the wider genus of adaptive processes. Adaptive processes come in the form of a coupling typically described in mathematical terms; something that has been called the 'agent-environment coupling'. In the context of perceptual processes, which are a special kind of adaptation, this coupling is called the 'sensorimotor loop'. A refined notion of adaptation (in the first, broader sense) was developed later in the enactivist framework for clarifying how all these concepts are interrelated (Di Paolo, 2005).

Given the former definitions, we can broadly distinguish between two varieties of enactivism. First, those that endorse the biological notion of agency as a self-sustaining system along with the idea that perception is based on a sensorimotor loop. Second, those that are only committed to this way of explaining perception and do not hold to the theory of biological agency that defines the first group. Among the latter authors we can find Noë (2005) and O'Regan (2012). Conversely, among the former theoreticians we can find Maturana and Varela (1987), Jonas (1966) and others. We will focus on the notion of 'normativity' provided by the latter group, and for this purpose we will analyze the most recent definition of this phenomenon: the one provided by Barandiaran et al. (2009).

This approach to normativity is given in terms of its co-emergence with individuality and action, and these three notions work as different conditions for agency. Also, this agency is at the service of the autonomy of the system. So, given the fact that living systems are autonomous or self-sustained, for enactive philosophers a prior assumption is required in order to understand how agency emerges. For the enactivists, the difference between machines and living beings is that for machines "no *intrinsic force or process* is lumping the components together, nor has the system as a whole (independently of us) a specific way of functioning and

demarcating itself from the rest" (Barandiaran et al., 2009; emphasis added). The individuality of the entity is not projected but recognized when we deal with living beings. The entity itself displays to us its own criterion of demarcation from the environment: an agent is then "a system capable of defining its own identity as an individual and thus distinguishing itself from its surroundings; in doing so, it defines an environment in which it carries out its actions" (Barandiaran et al., 2009, p. 3; the quotation appears in italics in the original).

But a question remains unanswered: what is the particular process (or force, as it is quoted above) that allows for this demarcation? The answer is nothing but its own actions: "agents define themselves as individuals as an ongoing endeavor and *through the actions they generate*" (Barandiaran et al., 2009, p. 3, emphasis added). So, agents, by means of their acting, demarcate themselves as independent entities — and in doing so, they define themselves and also define the environment in terms of exclusion.

Here we arrive at a crucial point: if the definition of individuality comes by means the agent's actions, how do we differentiate mere random movements from genuine actions? When we talk of an action-perception loop in order to describe the interaction of the agent with the environment we cannot consider that both parts are equally active in the interaction. The coupling of a leaf flowing in the air establishes a symmetrical relation: the weight of the leaf and the force of the wind regulate the process with the same degree of implication. As we have seen, living beings are different: they act upon the environment and thus they can be demarcated from it. This is an asymmetrical relation. Agents and environments do not play the same role in the coupling. Hence the sensorimotor loop (or any other coupling) is not like the leaf-air coupling: it is something provoked by the agent. Actions are not random because the agent tries to achieve a certain goal with them. An action is a goal-directed, normative movement. That purposiveness demarcates actions from other movements. Thus, "*agents have goals or norms according to which they are acting*, providing a sort of reference condition, so that the interactive modulation is carried out in relation to this condition" (Barandiaran et al., 2009; emphasis added). The statement quoted above is highly revealing: goals and norms are used interchangeably (Barandiaran et al., 2009, p. 5, footnote 2), and these norms are the reference condition by which we can say that agents are acting. Furthermore, the coupling with the environment is carried out in relation to this normative character that specifies the kind of interactions with the environment that are defined as 'actions'.

But what is a 'norm' from this enactive perspective? Is it a statement or an explicit rule like 'the queen can move any number of vacant squares horizontally, diagonally or vertically' or 'do not feed the animals'? Clearly not. First, no linguistic competence is necessarily involved in the employment of this kind of norm, and nor is interaction with other agents required. Rather, it seems that some process is normative when it establishes and maintains the individuality or self-sustenance of the system: "self production is a process that defines a unity *and a norm*: to keep the unity going and distinct" (Di Paolo, 2005, p. 434; emphasis added). A process

that benefits adaptation is a "norm given by self construction" (Di Paolo, 2009, p. 50).

Now we have the whole picture of enactivist agency: agents are systems that individuate themselves from the environment by means of their actions, and those actions are described normatively. We can talk of a *co-emergence of individuality and normativity*: even though enactivists separate these as different conditions, they also explain in what sense the two concepts are co-extensive or interrelated. It can be useful to briefly return to the quotes cited above: "agents define themselves as individuals as an ongoing endeavor and *through the actions they generate*" and "*agents have goals or norms according to which they are acting*". This amounts to saying that individuality is defined in terms of action and action is defined in terms of normativity. So, it is this "*deep circularity* and *entanglement* between networked processes, the self-maintaining conditions they generate and the interactions that the system establishes with the environment what [*sic*] makes agents so challenging to model and understand" (Barandiaran et al., 2009. p. 8; emphasis added). Recently, Barandiaran and Egbert (in press) modeled the normative behaviour of a unicellular agent based on these criteria. In their model they differentiated between derived and intrinsic normativity, and they claim that the latter is a central feature of living beings, which are able to establish and follow their own norms in order to keep up the self-sustainability of their structure and their ability to adapt to their environments. That is why, for these authors, enactivism is a new paradigm: because it establishes a theory of agency through which we can understand cognition and, specifically, the normative aspect of it.

Dispositions and norms

The first set of examples that could clarify the notion of 'normativity' defined by enactivism comes from the philosophical discussion of dispositions. Several authors have previously appealed to dispositions in order to explain the behaviour of physical objects but also of biological or rational agents (Ryle, 1949; Molnar, 2004; Mumford and Anjum, 2011). We say that sugar has the disposition to dissolve when put into water, neurons have the disposition to open their sodium channels when they receive a stimuli, and humans have the disposition to laugh when they listen to a joke. For some authors, these dispositional properties are defined as intrinsic, first-order, and real properties of agents and objects (Molnar, 2009). By 'first-order' we mean that dispositions are properties instantiated in individuals. The claim is that these dispositional properties are intrinsic to their bearers because they do not depend on the existence of any other object. Given these features, we can say for example that the fragility of a piece of glass is a property instantiated in a particular item, and also that the existence of that property does not depend on the existence of any object other than the piece of glass. The realist commitment to the property comes with the conclusion that, given its individuality and intrinsicality, a disposition does not need to manifest itself in order to prove it existence. Glass can maintain the property of being fragile even when it never breaks. We do not need the continuous manifestation of a dispositional property in order to assume its existence. Based on this, a special feature of dispositions is their

directedness: dispositions are directed to their reciprocal dispositional partners rather than to their manifestations. These reciprocal partners are other elements of the same kind that, given the right circumstances, play the role of triggering the manifestation of the property. For example, imagine a sugar cube: a sugar cube has the property of being soluble even when it is not able to show its solubility (let's say, even when it is covered by plastic wrap while submerged in a glass of water). In this case, the disposition exists even when it is not manifested. Also, following Martin (2008), we can imagine that some chemical product A has the property of being soluble when mixed with another chemical product B even if product B does not exist in the universe (e.g., product B has never been synthesized because it would be too expensive to do so). In any event, our realist intuitions towards dispositions incline us to consider that the product A has the property of solubility even if it will never be manifested.

An interesting feature of dispositions is that, applied to biological agents, they cover abilities and natural reactions as well as learned and innate responses. We can say that a dog once had the disposition to growl when the master picked up its bowl, but now it has the disposition to sit down when the master does the same thing. Also, given all the features mentioned above, dispositions can be useful for explaining interaction with the environment in a non-representational way: from a dispositional perspective, the basic unit of analysis is not the agent, but the agent-environment coalition. Dispositions, thus, explain the expected behaviours of certain agents under specific conditions.

Given this account of dispositions, we think that these properties are very useful for explaining different processes and behaviours of living agents, which are highly context-dependent. Thus, we are going to provide an example by which we can differentiate between a dispositional state and a normative behaviour: this will help us to show how the difference between following a rule and manifesting a disposition is blurred in the enactive account of perception. We will conclude that the subsumption of the dispositional within the normative is not helpful for explaining the different cognitive states of living agents. Imagine this situation: Manolo is a heavy smoker. To say that someone is a smoker is to make a dispositional attribution. Manolo is a smoker even when he isn't smoking. Being a smoker is being disposed to smoke a cigarette in certain circumstances (being a heavy smoker is to be so disposed in most circumstances). For instance, Manolo has the disposition to light a cigarette every time he sees one, as if an internal force pushed him to do it. One day Manolo goes to the doctor and he is told that if he continues to smoke, he is very likely to develop a chronic respiratory disease. He realizes that he *must* stop smoking. This realization did not cancel his *disposition* to smoke, at least not in the short term. Nevertheless, it did stop him from *manifesting* the disposition. What was it that stopped a deep-seated disposition from manifesting itself? The answer is clear: a *norm*. A norm can inhibit the triggering of certain dispositions even when the circumstances are otherwise entirely suitable for the disposition to manifest itself. From this perspective, a norm can inhibit but it can also eliminate a disposition after a certain number of corrections. So, the first

difference between a disposition and a norm is that the first is intrinsic and internal, but the second is not. This is why the internal force persists even when the agent follows a rule imposed by another agent (such as the smoker and the doctor). Given these features we can understand now why a norm, then, is different from a disposition. It is clear when we look at examples involving humans, but can this notion of 'norm' be applied to neurons or bacteria? It seems that there is no room for this conceptual tool in the explanation of the behaviour of unicellular agents. *Who* can correct the cell for not behaving in a certain way? How can an intrinsic disposition of a bacterium be inhibited by a norm, if the only dispositions that we can find in bacteria are those that allow for its survival (or, at least, that allowed for the survival of its ancestors)? It seems that there is something in the context of bacteria that is missing if we want to apply the concept of 'norm' to them.

Wittgensteinian norms

As we've seen, enactivists do not differentiate between *biological* dispositions and norms, and they label all these different processes 'normative'. This would be a minor problem if it were only a terminological issue. However, we think that the problem is also conceptual and ontological: what we have here is a disagreement regarding what a norm *is*, *how* the concept works and *in which context* it can be applied. We have seen that some enactivists claim that a solitary agent can both establish and follow its own norm. Now the question is whether that claim is acceptable. Is it intelligible to think of an agent who is able to establish and follow its own rules in isolation?

We think that Wittgenstein's discussion on rule-following is still very relevant to this question. In the well-known sections of his *Philosophical Investigations* devoted to this issue, he offers a battery of arguments to show that the answer should be negative (Wittgenstein, 1953, §§ 185-242). When we talk of following and establishing our own norm we are talking of establishing and following a special course of action. Wittgenstein wants us to imagine a situation in which somebody is teaching a pupil to count in a certain way and the teacher wonders why, after many repetitions, the student is still not doing it correctly. The first explanation of the student's behaviour is always to appeal to his natural reactions, to his natural inclinations, for answering one way rather than another. This suggests that we can distinguish between acting according to one's natural dispositions and acting correctly — acting according to a rule. So, following a rule seems to be something much more complex than naturally reacting. If all there is to following a rule was to act according to one's brute inclinations, then there would be no situation where learning could be thought to be necessary to coming to act *in the right way*.

If equating norm-following with acting according to one's unlearned natural dispositions is problematic, perhaps the enactivist, in her defense of the idea of biological norms, could appeal to the notion of interpretation. A sphere would be normative inasmuch as its inhabitants were capable of interpreting norms in such a way that their action was a case of following the rule under their interpretation. When discussing whether acting according to a norm can be

understood as offering an interpretation such that the action becomes subsumable by the rule, Wittgenstein comes back to the example of the pupil learning mathematics. After some successful exercises that seemed to show that he had mastered the use of the "+" sign (all involving numbers smaller than 1000), the teacher asks him "how much is 1000 + 2?" The student answers "1004". When the teacher tells him that this is not the right answer, he defends himself claiming that he is doing exactly what he was told: "I did as before. Wasn't the rule: add 2 up to 1000, 4 up to 2000, 6 up to 3000 and so on?" The student has managed to provide an interpretation of the rule behind the use of the "+" sign that covers all possible uses of the sign and is consistent with all of the examples he was exposed to during his learning. It is tempting to say that the pupil can act in accordance with his own criterion. At the very least, he seems to show a personal and systematic way to face stimuli after a number of repetitions and encounters with them. A defender of the idea of non-social, natural norms would argue that the habitual answer to the stimulus can become a norm (i.e., a well-established causal connection). In fact, on what else would the rule-following of an isolated agent depend than on its personal interpretation of the norm (i.e., on its own systematic way of reacting to a given stumuli)?

This way of understanding normativity seems deeply paradoxical: if acting according to a rule is no more than interpreting the rule in such a way that the action falls under it, then every action can be made out to accord with some interpretation of the rule and every action can also be made to conflict with an interpretation of the rule. Then there would be neither accord nor conflict here. If every idiosyncratic interpretation of the rule is right, then how we could say that somebody is wrong? It would seem that the concepts 'right' or 'wrong', which are tightly connected to the concept of 'norm', are of no use here. So, if everything is a norm, then nothing is a norm at all because nobody could distinguish what is normative from what is not. As Wittgenstein claims in § 201: "What this shows is that there is a way of grasping a rule which is not an interpretation, but which is exhibited in what we call 'obeying the rule' and 'going against it' in actual cases."

Our claim in this paper is that the idea of a bacterium establishing and following norms is just as problematic as the idea that all there is to grasping a rule is to behave in a way that coheres with some possible interpretation of the rule. To talk about norms is to talk about the possibility of being right and wrong, and this in turns demands that the agent be capable of distinguishing between "it is correct" and "it seems correct to me". Could anyone make such a distinction without having being corrected in the past? We believe not. Given the fact that the aspirant to being a rule-follower cannot be its own corrector, we claim that rule-establishing and rule-following need to be defined as a socially-mediated phenomenon. Norms can only emerge within a social context; norms are, then, social institutions. Norm-establishing is a social process. That is precisely why norms are external: because the criteria of correctness are shared across a community of agents.

The alternative, solipsistic conception of rule following makes following a rule analogous to speaking a private language (i.e., to following private, internal linguistic norms).

Norms must be guided by certain criteria that determine the correctness of their own applicability. These criteria are external in the sense that a single agent acting alone cannot establish them: if that were the case, senseless situations like the one discussed above would be common. But why is that situation 'senseless'? Because if somebody follows a rule and she cannot distinguish between following it and not following it, she cannot guarantee that she is following the rule in the right way. This point is explicitly stated by Wittgenstein: "Hence it is not possible to obey a rule 'privately': otherwise thinking one was obeying a rule would be the same thing as obeying it" (Wittgenstein, 1953, § 202). So, where do rules or norms come from? Don't they come from the agents that establish them? Sure, but this is not the same as saying that a solitary agent could be involved in the process of establishing a rule. Norms can only emerge within a social context.

Rule-*establishing* cannot be a private exercise, but what about norm-*following*? The enactivist claims: "even if the origin of some norms does not fully lie within the individual (e.g., social norms) it is always the individual who internalizes them" (Barandiaran et al., 2009, p. 6). What sort of process could this internalization be? For the purpose of answering this question, let's rescue another classic example from Wittgenstein, that of the beetle in the box (Wittgenstein, 1953, § 293). Let's assume that everyone in their own case knows how to follow a rule because they have internalized it. Each of us would walk around carrying a box and calling what is inside 'a beetle' — or, better, 'a norm'. Nobody can see inside anyone else's box, and everyone knows what a beetle (or a norm) is only through looking inside their own boxes. On the other hand, we all know how to use the concept 'beetle' or 'norm'. Suppose that in fact we all have different things in our boxes (or even imagine that there is nothing at all in them). The key point here is that the object in the box plays no role at all in our understanding of how to use the concept. In the same vein, we do not need to look inside us or appeal to any inner state to know what following a rule is. The criteria are *outside* the individual; they are located in the social community. They are *shared*. But they are not *objects*. We do not need to look for them as if they were part of our internal machinery. This is why norms are not individual-internal, but social-external processes, both when they are established and when they are followed. Being goal-oriented and having conditions of success and failure is necessary but not sufficient for being normative. Normativity also demands awareness of the possibility of error, training, habit and social learning.

Conclusions and further work

In this paper we have offered three inter-related arguments against enactivism's insistence on talking about norms at the level of simple, non-social agents: (1) the co-emergence of individuality and normativity is just taken for granted because the claim that these are mutually supporting ideas is viciously circular. A robust notion of agency related to an evolutionary history of adaptation and selection is sufficient to account for the singularity of living systems. (2) A notion of normativity as vague as the one offered by the enactive theory blurs the distinction between dispositional, individual, intrinsic natural processes and the social, external and institutional ones that

can inhibit the former group. A clear separation between phenomena defined by mere conditions of success and failure and phenomena characterized by correctness conditions makes explicit such a distinction. (3) The concept 'norm' can only be applied to what enactivists call 'social norms': the Wittgensteinian discussion of rule-following shows that there is something deeply paradoxical in thinking of the behaviour of an agent considered in isolation as being governed by norms. As usual, we should not be led to confusion by etymology: full-blown normativity and self-regulated behaviour are to be distinguished. A single agent cannot establish a rule because it is acting according to its dispositions; acting according to one's dispositions and acting according to a rule are not the same thing; and, finally, there is no need to appeal to any internalization of the norm to explain how agents follow them.

The behaviour of a cell is manifestly suitable for explanation in dispositional terms, because it cannot be divorced from its environment. We can say that its behaviour is rich enough to qualify it as 'goal-directed'. But it is not normative because there is no socially-established norm that could inhibit any of the cell's intrinsic dispositions. Inasmuch as the criteria of correctness of that hypothetical norm are not shared, the cell could not possibly distinguish between instances when it is acting according to the norm and instances when it is not. Neither could we: the distinction between failure due to the cell's behaviour and failure due to, say, a hostile environment cannot be made for actions that are mere manifestations of dispositions. Our claim is that a cell's behaviour may be insufficient to guarantee its survival in some environment, but that such failure does not entitle us to consider the behaviour incorrect. This is why the only wrong cell is the dead one.

Our qualms with enactivism's excessively liberal use of normative considerations is no obstacle to our sympathy with enactivism's anti-representationalist commitments, as well as its emphasis on embodiment, situatedness, the active character of perception, and the centrality of the agent as a whole. We also agree that perceptual relations with the environment can be explained by means of looping processes. However, we think that this anti-representational approach to cognition is better developed by ecological psychology (Gibson, 1979). Gibsonians gave an account of perception in a way that is much more externalist, biosemiotic, and structure-independent than the sensorimotor contingencies defended by the enactive view. This is the reason why ecological psychology has provided a better account of learning than enactivism, even though they start with the same anti-representational assumptions (Jacobs and Michaels, 2007). Enactivism is too closely focused on the internal structure of the organism and concedes too little attention to the explanatory role that the environment plays with respect to perception and action. We also depart from the enactivists regarding their faith in autopoiesis being the best explanation of every aspect of biological processes: from the emergence of agency to the emergence of perception. We also do not see autopoiesis as being the best explanatory framework for processes such as adaptation and cognition. In fact, as we have seen, the enactive, co-emergent explanation consists in subsuming all biological processes into just one: the recursive loop made by all systems of every organism. This may well be the best

answer to the question of how all organisms are able to maintain their stability through time, but that does not amount to defining agency, adaptation, cognition and the rest of the set of biological processes at once. Take the example of adaptation, a process the enactivists sometimes call 'adaptivity' (Maturana and Varela, 1984; Di Paolo, 2005). This process is based on the recursive loop we have mentioned, and we can apply the same logic of recursivity to the relations of the agent with the environment: an adaptive capacity is one that is able to regulate its relation with the environment in order to keep the agent within a state of viability. Organisms can detect tendencies in which the agent approaches (or recedes from) the boundary of viability. As any biologist would concede, this formulation is insufficient to account for adaptation in the full sense. Adaptation is a trait that contributes to the fitness and survival of individuals but it also needs to be explained as the result of processes of natural selection, and reference eventually needs to be made to species and populations (Darwin, 1859; Huxley, 1942; Williams, 1966; Mayr, 1983). If we want to give a full account of why an agent is adapted, we necessarily need to appeal to its evolutionary history and talk about how natural selection works. This is a question answered only at the macroscopic level and by means of reverse engineering (Dennett, 1995), not by looking at the looping processes of individual agents. We think that the excessive emphasis on the logic of looping processes is leading enactivism towards an underestimation of natural selection, the role of populations, and the different levels of explanation involved. Not all questions in biology are answered by redirecting the answer to the looping processes of self-sustenance of individual agents. Some questions (why a trait has evolved this way rather than that way, *why* we have perceptual system at all, etc.) are answered by appealing to the supra-agential realm and this means by appealing to how natural selection works. Whereas other questions (*how* do we perceive, etc.) are answered appealing to looping processes (ecological, perception-action loops). There are different questions addressed by different levels of explanation. We do not think, as enactive theorists seem to endorse, that all biological processes can be explained by means of their looping and co-emergent logic and by appealing to the autopoiesis of individuals.

We think that this philosophical discussion is clearly of interest for computer scientists for two very important reasons: it is important not to confuse levels of analysis and also not to misattribute properties or predicates to agents that do not fulfill the right criteria of application. An unicellular agent cannot be wrong because there is no room for norms in its behaviour. For following a norm some conditions are needed: (1) a community, (2) the possibility to err, (3) correction criteria for the right application of a concept or a right way of behaving in a certain context, (4) the possibility to differentiate between following a norm and thinking that one is following a norm, (5) to be sanctioned by a community in order to understand which are these criteria and how to differentiate between what one thought she was doing when following a rule and what she was really doing. Even when a Wittgensteinian strategy is not committed to offer necessary and sufficient conditions for defining a concept, these previous points can summarize more or less some features that are common to any notion of 'norm'. We think that,

summarized in the previous points, a computer scientist does not offer anything new when he claims that a unicellular isolated agent can follow norms. That claim only shows that, even when he could design a really good model, the scientist never got the conceptual point of what a norm is. Even though it is surely possible to describe all of the different levels of agency from a naturalistic viewpoint (making use of our best empirical evidence from the biological sciences), introducing the most complex concepts, such as 'normativity', when studying the most primitive forms of agency is not a good strategy. A better strategy would be to focus on what have been called the 'major transitions in evolution' (Maynard Smith and Szathmáry, 1995); that is, to focus on the conditions under which new organizational levels appear rather than taking them for granted. Unlike enactivism, we reject the idea that the naturalization of normativity can be made by normativizing nature: all that is rational is real, but not all that is real is rational.

Acknowledgements

References

Barandiaran, X. E. and Egbert, M. D. (in press). Norm-establishing and norm-following in autonomous agency. To appear in *Artificial Life*.

Barandiaran, X.E. Di Paolo, E. and Rohde, M. (2009). Defining agency. Individuality, normativity, assymetry and spatio-temporality in action. *Journal of Adaptive Behavior*, 17 (4) 367-386 in Rohde, M. and Ikegami, T., editors, special issue on Agency in Natural and Artificial Systems.

Canguilhem, G. (1965). *La connaissance de la vie*. Vrin, Paris.

Darwin, C. (1859). *On the Origin of Species*. John Murray, London.

Dennett, D. C. (1995.) *Darwin's Dangerous Idea*. Simon and Schuster, New York.

Di Paolo, E. (2009). Overcoming autopoiesis: a enactive detour on the way from life to society. In Magalhaes, R. and Sanchez, R., editors, *Autopoiesis in organization and information systems*, pages 43-68. Elsevier.

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4: 429-452.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Houghton-Mifflin, Boston, MA.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton-Mifflin, Boston, MA.

Huxley, J. S. (1942). *Evolution: the Modern Synthesis*. Allen and Unwin, London.

Jacobs, D. and Michaels, C. (2007). Direct Learning. *Ecological Psychology*, 19: 321-349.

Jonas, H. (1968). Biological foundations of individuality. *International Philosophical Quarterly*, 8 (2): 231-251.

Martin, C. B. (2008). *Mind in Nature*. Oxford University Press, Oxford.

Maturana, H. and Varela, F. (1987). *The Tree of Knowledge*. Shambhala, Boston, MA.

Maynard Smith, J. and Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford University Press, Oxford.

Mayr, E. (1983). How to carry out the adaptationist program? *The American Naturalist*, 121 (3): 324-334.

Mumford, S. and Anjum, R. (2011). *Getting Causes from Powers*. Oxford University Press, Oxford.

Molnar, G. (2004). *Powers: a Study in Metaphysics*. Oxford University Press, Oxford.

Noë, A. (2005). *Action in Perception*. The MIT Press, Cambrige, MA.

O'Regan, J. K. (2012). How to build a robot that is conscious and feels. *Minds and Machines*, 22 (2): 127-136.

Ryle, G. (1949). *The Concept of Mind*. Hutchinson, London.

Stewart, J. Gapenne, O. and Di Paolo, E. editors (2010). *Enaction: Toward a New Paradigm for Cognitive Science*. The MIT Press, Cambridge, MA.

Williams, G. C. (1966). *Adaptation and Natural Selection*. Princeton University Press, Princeton, NJ.

Wittgenstein, L. (1953/2001). *Philosophical Investigations*. Blackwell, Oxford.