Cybernetics And Human Knowing. Vol. 15, nos. 3-4, pp. 50-56

Meaning Needs Functional Emergence:
Shedding Novel Light onto Difficult Problems

Argyris Arnellos, Thomas Spyrou, John Darzentas[1]

1 Introduction

We would like to begin by thanking Gonzalez and Pessoa (this issue) for their interesting commentary. Of course, a critical commentary always needs the adoption of a perspective, and an improper perspective could lead to mistakes and misunderstandings. In this paper we will attempt to provide some explanations in order to better clarify our theoretical theses and overcome those misunderstandings.

Gonzalez and Pessoa's main argument is that "from the very way that the concepts are defined, questions concerning the possibility of building an autonomous system would receive a negative answer" (this issue, p. 43), with respect to whether robots or any artificial system can be considered as an autonomous agent. In the light of this perspective, they conclude that this is a simple consequence of the definitions adopted and as such, little novelty is added to the already existent research on autonomous and cognitive robotics and ALife.

A thorough reply to their remarks would probably be too long for the purposes of a paper, but let's start by trying to indicate the basis of their commentary.

2 Agency is a Systemic Capacity

In Arnellos, Spyrou, Darzentas (in press) it is made clear that agency is a systemic capacity. Moreover the relevant notions/properties of an agent cannot just be added together in an arbitrary manner. As such, the concepts of autonomy, functionality, intentionality and meaning do play a fundamental role in the characterization of a natural agent, but these notions are unjustifiable on their own. In the same way, interactivity is also a fundamental notion, but truly useless and difficult to be handled (either theoretically or experimentally) on its own. One would think that the notion of goal, goal-orientation and purpose in general would solve, at least the theoretical problems, but again, where is this goal or purpose coming from and where does it go?

Then, one is willing to dig in another resolution and to mention issues of self-organisation, self-reference, cohesion, organizational, process and interaction closure, normative functionality and the respective representations, and so forth. But still, the problem of how to handle these difficult and philosophically loaded notions has not so far attempted to be answered.

---

1. Department of Product and Systems Design Engineering, University of the Aegean, 84100, Syros, Greece. Email: arar@aegean.gr, tsp@aegean.gr, idarz@aegean.gr

It is this point that Gonzalez and Pessoa seem to neglect in their commentary. The work of Arnellos, Spyrou, Darzentas (this issue) is exactly on this track, as it is an attempt to provide a framework where all these theoretical notions can find a justifiable way to be integrated so that their emergence is naturalistically acceptable. Moreover, the second part of Arnellos, Spyrou, Darzentas (this issue) work tries to indicate and map these properties and their explicit or implicit consequences in the certain technical solutions being applied in the design of artificial agents.

Having said that and keeping in mind that we are aiming at a naturalized analysis of agency and of the emergence of agential capacities, it is very difficult that our thesis will not be shown as being prejudiced regarding a negative answer on the autonomy of contemporary artificial agents. However, the goal of Arnellos, Spyrou, Darzentas' work (this issue) is to provide a theoretical framework regarding the functional emergence of autonomy in an agent, to detect the gap between theoretical and practical aspects of analyzing and building autonomous agents and to provide some possible design guidelines.

As it was expected due to the theoretical emphasis of the paper, Gonzalez and Pessoa are criticizing whether the theoretical framework is essential for the arguments that are being developed. Certain disagreements are mentioned in their commentary. We will discuss each in turn.

3 Normativity via Functional Norms Precedes Intentional Behavior

Gonzalez and Pessoa argue that in our theoretical framework "there seems to be no interest in explaining how teleological or intentional behavior can arise in previously non-teleological and non-intentional systems" (this issue, p. 45). In this case, the explanation for the emergence of normative function seems circular.

But this is not the case. In section 2.2 we specifically mention that the emergence of cohesion in a self-organising system is primarily solely due to its organizational (functional) dynamics. This cohesion, at least at this fundamental level, can only be explained with respect to the causal roles of the constituents and the relations among them. However, as it is also suggested in section 2.2, there are certain kinds of cohesion, with different types of correlations between the respective functional processes, and not all of this kinds of cohesion will provide a genuine autonomy for a system. One of these cases is the case of self-maintenant systems, which do not need to be living systems, but just non-living self-maintaining far-from-equilibrium systems, such as a candle, a flame, a tornado, and so forth. As Vehkavaara (2003, pp. 565-566) soundly states,

> these systems are serving their self-interest and, as a consequence, are 'staying alive' (without being living), they are nevertheless not *trying* to serve it. Their self-interest is not *forcing* or 'suggesting' them to do anything. They are not *seeking* how they can survive, but they just *happen* to have such a structure that fulfills their sole self-interest and existential precondition for some period. There self-maintenance does not yet give birth to any real growth or increase in complexity." (Vehkavaara, 2003, pp. 565-566, emphasis in the original).

Vehkavaara's suggestion is in accordance with our argument that for a self-maintaining system this is the phase where constructive processes dominate the system through process closure. The result is a functionality which is unbreakably related to the maintenance of the systemic cohesion and in consequence, of its self-organisational dynamics. This is not a genuinely autonomous system, it is non-teleological and non-intentional and it is neither living nor representative one. However, this asymmetric functionality (between system and environment), which results in unintentional self-maintenance provides functional norms (i.e. primitive/fundamental normativity), which forms the basis for the emergence of other more developed normative functions in the representative/interpretive phase of a system. But this organizational level will not come until life comes about. As it is thoroughly explained in section 2.3, this will not happen until the unintentional self-constructive processes are complemented with interactive processes forming an interpretive asymmetry, which also provides the capacity for "meaningful critique regarding the functional and the dysfunctional with respect to the maintenance of the system" (this issue, p. 26). This capacity is a normative one and such systems are truly autonomous. In this perspective, that case where "normative functions emerge as a contribution for the autonomy of the agent, and with the goal of satisfying the respective functional norms" (this issue, p. 27) cannot be considered as circular, but as vital for the maintenance and further enhancement of the autonomy of the system.

We are closing this remark by noting that the aspects of this section are highly related to the notions of *purpose* and to certain aspects of the *materiality* of the system under consideration. These aspects will be discussed in following sections.

4 Interaction Based on Thermodynamics Leaves no Room for non-Pragmatical Considerations

Gonzalez and Pessoa agree that all the features we suggest in Arnellos, Spyrou, Darzentas (this issue) are essential to agency, but they are wondering if the emergence of such properties can be explained by a system's theoretic framework of second-order cybernetics, while on the same time, they are posing questions related to whether our view is a constructivistic or a realistic one.

This is a considerable remark by Gonzalez and Pessoa, which is difficult to be answered if one decides to stay merely in the constructivist camp. However, we have adopted a different view, where interaction complements the constructive dimension providing interpretive constitution to our theoretical framework (Arnellos, Spyrou, Darzentas, in press). That said, we have no problem translating from a constructivist language to a realist one, since in accordance with our theoretic framework, a nervous system organizes itself so that it computes a stable representation of reality as this representation emerges in the interaction of the system with its environment.

In this perspective, it seems that it cannot exist as a merely constructivist or realist view, since any construction should be able to be tested in the system and by the system itself, the possible representational error should be functionally available in the

system itself (Bickhard, 1993, 2000; Arnellos, Spyrou, Darzentas, in press) and the test will take place at the system's interaction with the environment. In other words, whatever the status of the external environment, the results of the interaction should be internally and functionally available.

However, Gonzalez and Pessoa continue their criticism by asking how we come to attribute autonomous self-organization and autopoiesis only to living agents, while we are also presenting the double closure of the sensorimotor system as a case of emergence and downward causation. First of all, this is not an artificial network, but a part of a greater living network of organizational processes, which cannot exist outside of it. This is the reason why Ziemke and Thieme's (2002) model cannot emerge on its own and it does not scale to greater levels of complexity. Again, Gonzalez and Pessoa ask about the means with which the kind of features being supported by such organizational networks may arise in phylogenesis and they are also asking why the artificial network suggested by Ziemke and Thieme's (2002) does not fulfill the characteristics for self-organization indicated by Collier (2004a).

We will try to answer these two comments using as a basis the criticism from Gonzalez and Pessoa on (this issue, p. 45), stating that our theoretical framework gives no emphasis to the concepts of *natural selection* and *noise*, as two essential concepts regarding the evolution of self-organising systems. Well, natural selection has many problems related to normative functionality (see e.g., Collier, 2004b), as this is defined and described in Arnellos, Spyrou, Darzentas (in press, this issue). The main problem is that in its abstract and disembodied version seems to be a possible solution to the problems discussed in the present paper, but once studied in its embodied version, the picture seems to turn upside down. Specifically, natural selection is not an abstract/formal process but an energetic/thermodynamic one. Its thermodynamic aspect is being taken care by the constructive/interactive processes of the system itself and as such, each noise triggering the self-organisation of the system is integrated into the respective functional processes, sometimes successfully and sometimes not. In either case, the phylogenetic aspects of such evolution are bounded to the materiality of the system and as Deacon (2006) suggests, in living systems selection is directly depended on the energetical aspects of the respective materiality.

Of course, the exact mechanisms are still unknown, but we have tried to provide some hints towards this direction by suggesting some logical features and conditions based on which such evolution could be started or even better, based on which a system can be said that it evolves while its interaction with the environment. However, the problem of the materiality of the system that would be able to support such features (process, interaction closure, formation of asymmetries that will result in the internal construction of new functions, etc.) remains. Nevertheless, we think that researchers of A-life should consider these features and conditions while trying to build systems that will develop intentions out of non-intentional dynamics.

5 Meaning is Emergent and has a Functional Substratum

Gonzalez and Pessoa agree that meaning should have a functional substratum and that the respective functionality should be emergent in the interaction with the environment. But they seem to strongly criticize the suggested framework as being mixed up with problematic *internalist representational* conjectures (this issue, p. 47). Well, we think that this is due to a misunderstanding due to their missing out the descriptive and explanatory power of the notion of interaction, as well as of the feature of normativity. Let us start from the latter.

Although we do say that "internal productive interrelations acquire a cohesive functional meaning in a collective way, since they contribute to the overall maintenance of the system" (this issue, p. 23), we do not adopt a pansemiotic position. As it is thoroughly explained in Arnellos, Spyrou, Darzentas, meaning is expressed through the choice of a certain function in the light of several possible choices and with respect to a purpose of the system. In this perspective, any cohesive system at the level of self-maintenance cannot be said to express meaning other than the one defined at the level of its respective functional norms (see section 3). This should be a satisfying answer to what happens in "simple" self-organizing systems that do not have a nervous system.

And now let's turn to the issue of the *internalist representational* conjectures. Bickhard's interactivist model, which we have tried to integrate in the suggested framework regards recursively self-maintaining systems – systems with more than one function at their disposal and systems which have already a functional substratum – as exhibiting certain functional norms. In general, each interaction of such systems with their environments will result in a specific internal outcome for the set of subsystems engaging in the specific interaction. We have specifically stated in section 2.4 that these outcomes depend both on the functional organization of the subsystems and on the environment. These outcomes create a differentiation in a way that they predicate the existence of a certain type of environment. So, either constructively or realistically, this differentiation is the only epistemic contact of the system with its environment.

But this differentiation does not provide any representational content as it is not considered as a representation in the suggested framework. So, there is not any internalistic representational conjecture. On the contrary, what we really have at this case is a functional state of the respective subsystems which is evolutionary connected to other functional processes of the system. These processes are forming a functional cohesion since we consider a recursively self-maintaining system. According to what has been mentioned in section 3 (this paper) this integrated functionality serves some purposes, in spite of their unintentional basis. These purposes may just be of the kind with which the system maintains itself, or as Bickhard (1993, 2000) suggests, it may be a simple goal system where a certain process, whose conditions of operation are driven by the environment, may direct the flow of operation to another process, which may either redirect the flow to the former process or outside the system. This is the simplest type of a goal-oriented system and it can be found in living and in non-living

systems. What needs to be clearly understood is that the possibility of this redirection of operation between these two processes, or between many more processes in more evolved systems creates a *representational content* which emerges in the system's interaction with the environment.

The representations that will emerge depend both on the system and on the environment. Moreover, these representations should be considered as the anticipation of the system regarding its interactive capabilities towards the respective environment. As the system develops new representations, it further develops new anticipations and hence, normative functionality emerges. Therefore, it is totally wrong and misleading to consider a continuous modulation of action, taking place in a context of downward causation and based on any interactive context, while being driven by the system's norms, as an internalist representation.

At this point one should note that in the pragmatic context, a representation could be wrong. This is true and at this point Gonzalez and Pessoa are right to argue in favor of the adoption of mechanisms of learning. We have deliberately leaved learning out of the picture, as it was implicitly assumed when we mentioned that the interactive capabilities of the system, that is its anticipation may be inappropriate and this is an error which should be detectable by the system itself. So, we agree with Gonzalez and Pessoa that learning is necessary, as we have extensively mention elsewhere (Arnellos, Spyrou, & Darzentas, 2007, in press). But there can be no learning unless the system has a functional organization where any representational error would be internally detected and also available to the system itself. In this perspective, any kind of learning, (via communication with others, training, etc.) requires a capacity of constructing anticipations and representational content with a reference to a certain purpose. In other words, it requires a communication via meaning though the respective functional structures (Arnellos, Spyrou, & Darzentas, 2007).

At this point we are ready to answer to the last question posed by Gonzalez and Pessoa, namely whether "artificial systems could evolve in such a way that they would autonomously acquire learning mechanisms with indicators of relevant information in their selective interaction with the external world?" (this issue, p. 48). As we mentioned in section 3, the emergence of normativity takes place on the basis of primitive functional norms, which may even correspond to the fundamental purpose of self-maintenance, which in turn may be purely unintentional and based solely on the properties of the physical system under consideration. Nevertheless, according to the suggested framework, this is the functional norm upon which the emergence of other more developed norms will take place. As such, it seems that the condition to be satisfied is the evolution of newer purposes on the basis of a purpose denoting *self-interest*, independently whether it is a living or a non-living system.

A characteristic example of such a system is the "Big-Dog" (http://www.bostondynamics.com/content/sec.php?section=BigDog). Although not the result of endogenous evolution, this can be considered as an artificial agent that serves the self-interest of not falling down. Is this enough in order to be characterized as autonomous? The answer is negative based on our framework, but its functional norm

could be the basis for the development of an autonomous agent if there could be a way of evolving emergent capacities which would be functionally integrated on this norm. Again, materiality is the key to this question.

6 Conclusions

There is a significant difference between primitive functional norms and any further developed normative functionality. The suggested theoretical framework aims in providing directions for research in AI and is not meant to be an absolute theory for building autonomous artificial systems capable of evolving new meanings based on new functions. As such, the conditions for its falsification could not be other than the arguments that will render the framework as being non-naturalised as possible. But we all know that this is the case by default. However, the commentary of Gonzalez and Pessoa has by no means been directed towards the indication of architectures that are trying to support at least one of the features/aspects being mentioned by our framework.

Therefore, AI does not move towards such a direction, but in our humble opinion, AI is tinkering at the moment and this is something that should make even more evident the power of an interactivist framework of building autonomous artificial agents, as well as the importance of materiality. All these aspects should probably make our judgment, regarding the possibility of the existence of autonomous artificial agents, much more easier, but we also think that the analysis of our theoretical framework combined with the evaluation of the several attempts so far (at the technical level), are shedding new light in this very difficult problem.

References

Arnellos, A., Spyrou, T., & Darzentas, J. (2007). Cybernetic embodiment and the role of autonomy in the design process. *Kybernetes, 36* (9/10), 1207-1224.

Arnellos, A., Spyrou, T., & Darzentas, J. (in press). Towards the naturalization of agency based on an interactivist account of autonomy. *New Ideas in Psychology.*

Bickhard, M. H. (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence, 5,* 285-333.

Bickhard, M. H. (2000). Autonomy, function, and representation. *Communication and Cognition — Artificial Intelligence, 17,* (3-4), 111-131.

Collier, J. (2004a). Fundamental properties of self-organization. In V. Arshinov & C. Fuchs (Eds.), *Causality, emergence, self-organisation* (pp. 150-166). Moscow: NIA-Piroda.

Collier, J. (2004b) Self-organisation, individuation and identity. *Revue Internationale de Philosophie, 59*, 151-172.

Deacon W. T. (2006) Reciprocal linkage between self-organizing processes is sufficient for self-reproduction and evolvability. *Biological Theory, 1* (2), 136-149.

Vehkavaara, T. (2003). Natural self-interest, interactive representation, and the emergence of objects and *Umwelt*: An outline of basic semiotic concepts for biosemiotics. *Sign Systems Studies, 31*(2), 547-587.

Ziemke, T. & Thieme, M. (2002) Neuromodulation of reactive sensorimotor mappings as a short-term memory mechanism in delayed response tasks. *Adaptive Behavior*, *10* (3/4), 175-199.