



# Cybernetic embodiment and the role of autonomy in the design process

The role of  
autonomy in the  
design process

1207

Argyris Arnellos, Thomas Spyrou and John Darzentas  
*Department of Product and Systems Design Engineering,  
University of the Aegean, Syros, Greece*

## Abstract

**Purpose** – This paper aims to develop the role of autonomy in the emergence of the design process. It shows how the design process is facilitated by autonomy, how autonomy is enhanced through the design process and how the emergence of anticipatory and future-oriented representational content in an autonomous cognitive system provides the functionality needed for the strengthening of both its autonomy and the design process, in which the autonomous cognitive system purposefully engages.

**Design/methodology/approach** – Initially, the essential characteristics of the design process and of the cognitive systems participating in it will be identified. Then, an attempt to demonstrate the ability of an enhanced second-order cybernetic framework to satisfy these characteristics will be made. Next, an analytic description of the design process under this framework is presented and the respective implications are critically discussed.

**Findings** – The role of autonomy is crucial for the design process, as it seems that autonomy is both the primary motive and the goal for a cognitive system to engage in a design process. A second-order cybernetic framework is suitable for the analysis of such a complex process, as long as both the constructive and the interactive aspects of a self-organising system are taken under consideration.

**Practical implications** – The modelling of the complex design process under the framework of second-order cybernetics and the indication of the fundamental characteristics of an autonomous cognitive system as well as their interrelations may provide useful insights in multiple levels, from the purely theoretical (i.e. better understanding of the design process and the conditions for each creative fostering), to the purely technical (i.e. the design of artificial agents with design capabilities).

**Originality/value** – The innovative aspect of the paper is that it attempts an analysis of the design process under a framework of second-order cybernetics, by attempting to analyse and explain the emergence of such a process from the point of view of an autonomous cognitive system. This results in some interesting implications regarding the nature of the design process, as well as regarding its “mechanisms” of emergence and evolution, with respect to the characteristics of the participating autonomous systems.

**Keywords** Autonomy, Second-order cybernetics, Design process, Functionality, Closure, Representational content, Anticipation, Interaction, Cybernetics, Design

**Paper type** Conceptual paper

## 1. Fundamentals of the design process

### 1.1 *Design needs autonomy and interactivity from its participants*

It is widely acknowledged that the task of defining design and analysing the design process is not something trivial. Banathy (1996, pp. 11-13) lists up to 24 design definitions. Although the definitions differ from each other, they appear to share a common opinion, namely, that design, in general, and the design process, in particular, is considered as a cognitive activity. For instance, this is clearly implied by Simon (1999, p. 111), when he states that “everyone designs who devises courses of action



---

aimed at changing existing situations into preferred ones.” In a more inclusive manner, Friedman (2003) argues that most definitions of design describe it as a goal-oriented process, where the goal is a solution to a problem, the improvement of a situation or the creation of something new and useful.

Therefore, given that the ability to act upon an environment in order to effect a goal-oriented attribution of a certain purpose belongs to a cognitive agent, *design should primarily be attributed to a cognitive agent and hence, it should have as its basis the cognitive process.*

On the other hand, Kampis (2002) suggests that a strong notion of agency calls for *interactivity*, that is, the ability of an agent/cognitive system to perceive and act upon its environment by taking the initiative; *intentionality*, the ability of an agent to effect a *goal-oriented* interaction by attributing purposes, beliefs and desires to its actions; and *autonomy*, which can be characterized as the ability of an agent to function/operate intentionally and interactively based only on its own resources. Furthermore, it seems that there is a very interesting interdependence between these three properties. Specifically, as Collier (1999) suggests, there is no *function* without *autonomy*, no *intentionality* without *function* and no *meaning* without *intentionality*. The interdependence is completed by considering *meaning* as a prerequisite for the maintenance of system’s *autonomy* during its purposeful *interaction* with the environment.

These properties and their interdependence are characteristics of the strong notion of agency (i.e. the one exhibited by living systems), which is considered as emergent in the functional organization of the living/cognitive system. The term “functional” is used here to denote the processes of the network of components that contribute to the autonomy of the cognitive system and particularly, to the maintenance of the system as a whole (Ruiz-Mirazo and Moreno, 2004). On the other hand, meaning, if it is not to be considered as an ascription of an observer, should be linked with the functional structures of the system. Hence, meaning should guide the constructive and interactive processes of the functional components of the system in such a way that these processes maintain and enhance its autonomy. In this perspective, the enhancement of autonomy places certain goals by the system itself and hence, the intentionality of the system is guiding its behavior through meaning. It should be noted that in such an autonomous system intentionality is not reducible to the processing of meanings, nor are the combinations of meanings bringing forth any “aboutness.” On the contrary, meaning and its functional substratum are the defining properties of an autonomous agent that may act intentionally. In other words, an autonomous system may act intentionally if its actions are mediated by meaning.

Hence, it appears that for a cognitive system (an agent) to be able to engage in a design process, it needs to exhibit the degree of autonomy that will provide for the functionality that is needed, in order to support its intentional and purposeful interaction with the environment, the result of which will create new meanings that will enhance its autonomy. Moreover, the design process has an interactive and a goal-oriented nature, which results from the interactivity and the intentionality of each cognitive system that engages in the design process.

So far, *the characteristic properties of a cognitive system able to engage in a design process* have been mentioned. However, the definition and an analysis of the design process cannot be solely based on the properties of the cognitive agent discussed above. As it will be shown in the next sections, there are also some characteristic

properties of the design process that should be supported from each autonomous cognitive system participating in the design process.

### *1.2 Ill-definedness and the open-ended nature of the design process*

The goal-oriented nature of the design process is usually related to a problem, or a set of problems, the nature of which is constitutive of the design process itself. Most design problems are defined in terms of properties and needs of the people who will use the outcome of the design process (an *artifact*, which can be material or immaterial), the purpose it has for them and the form the artifact should possess in order to be deemed successful. Such design problems are *ill-defined* and the possible solutions are not clear from the beginning. Design solutions are almost never predictable and there is never sufficient information to define the desirable goal state in advance. Particularly, finding a solution requires in addition finding out what the “real” problem is, which in respect to human-centred problems is impossible. The phases of solving a problem and specifying what the “real” problem is, are developing in parallel and drive each other. As Heylighen and Bouwen (1999) argue, solutions and problems co-evolve during the whole design process.

The ill-definedness of design problems is also considered by Banathy (1996) when he argues that design confronts interrelated complexes of problems. In particular, he states that design confronts “a system of problems rather than a collection of problems” (Banathy, 1996, p. 29), and he notes that “design problems are ill-structured and defy a straightforward analysis” (Banathy, 1996). Therefore, he adopts an evolutionary approach to design (Banathy, 1989, 1996, 1998, 2000; Laszlo, 2001), which seems to justify both the ill-defined and open-ended aspect of the design process. Particularly, Banathy argues strongly in favor of a design inquiry as the attempt to find out what should become real, in terms of discerning what would be a desirable addition to the real world.

Therefore, the design process seems to be considered as a form of inquiry driven by intentional action. Accordingly, the meanings of each cognitive system participating in the design process are continuously evolving and they are constantly incomplete and imprecise, no matter how much the problem solving progresses. Hence, design problems are also *open-ended*. There are different logical paths to reach a design solution, that is, different cognitive systems construct different meanings of the design problem and consequently, provide different meaning-based outcomes as a respective solution. This turns designing into a process which is difficult to model and even more difficult to prescribe.

### *1.3 The design process is future-anticipative*

The receiver (in general, the user in every instance of the design process) of an artifact will interact with it on the basis of his own individual experience. Considering that each user’s experience and hence meanings are different, the *content* of the design process, namely, that which is being conveyed, during the design process, by a cognitive system to the other cognitive systems engaging in the same design process, or/and that which is being conveyed, after the end of the design process, from the design system itself to one or more cognitive systems outside of the respective design system[1], should not be understood to be the artifact itself. On the contrary, the content is subjectively interpreted and changed by the user’s cognitive processes, while

in turn, he is purposefully engaging in future design processes. The different interpretation of content from multiple receivers with different meanings implies that the design process should have the potential to be directed towards many different possible outcomes and their consequences. In other words, the design process should have an *anticipatory* nature by which it will be placed in a pragmatic context and simultaneously be projected against the future, using different directions and time scales, (Banathy, 1996; Nadin, 2000; Jonas, 2001). It is this orientation towards the future that makes design different from mere problem-solving. Its interactive nature implies a new kind of anticipation for each cognitive system engaging in the design process, that learn from the past and appraise what is presently useful and desirable by simultaneously projecting their content into the future.

As it has already been noted, a complete description and analysis of the design process is not an easy task. The design process requires the engagement of individual cognitive systems in intentional and purposeful interactions with their environment and consequently with each other, in order to be able to fulfill their ill-defined goals. Such cognitive systems should have an autonomy that will guide them through this kind of interactions, based on their open-ended anticipative functionality. In other words, such cognitive systems require a certain type of embodiment.

In the next section it is argued that the design process and its characteristics call for a 2nd-order cybernetic epistemology. An analysis of the design process under the framework of second-order cybernetics is attempted, and it will be argued that *a second-order cybernetic system may exhibit the type of embodiment which forms the basis for a system to be able to engage in a design process.*

## **2. Second-order cybernetics and the design process**

### *2.1 Closure and self-reference for self-organisation*

Glanville (2001) suggests that the design process should primarily be examined within a cognitive framework based on 2nd-order cybernetic epistemology. In that case, a cognitive system is able to carry out the fundamental actions of distinction and observation. It observes its boundaries and it is thus differentiated from its environment. As the cognitive system is able to observe the distinctions it makes, it is able to refer the result of its actions back to itself. This makes it a *self-referential* system, providing it with the ability to create new distinctions (actions) based on previous ones, to judge its distinctions, and to increase its complexity by creating new meanings in order to interact (Luhmann, 1995). The self-referential loop can only exist in relation to an environment, but it also disregards the classical system-environment models, which hold that the external control of a cognitive system's adaptation to its environment is replaced by a model of systemic *closure* (von Foerster, 1981).

Owing to that closure, the self-reference of an observation creates *meaning* inside the cognitive system, which is used as a model for further observations in order to compensate for external complexity. The system which operates on meaning activates only internal functions and structures, which von Foerster (1981) calls *eigenvalues*, postulating some stable structures, which are maintained in the functions of the cognitive system's organisational dynamics (Rocha, 1996) and which serve as points of departure for further operations during its interaction with the environment. Indeed, this closure is functional in so far as the effects produced by the cognitive system are

the causes for the maintenance of its systemic equilibrium by forming new and more complex organisations.

With system closure, environmental complexity is based solely on system observations, thus, system reality is observation-based. As von Foerster (1976) argued, the results of an observation do not refer directly to the objects of the real world, but instead, they are the results of recurrent cognitive functions in the structural coupling between the cognitive system and the environment. In particular, von Foerster (1976, p. 266) states that “Ontologically, eigenvalues and objects, and likewise, ontogenetically, stable behavior and the manifestation of a subject’s ‘grasp’ of an object cannot be distinguished.” Thus, each new function based on observations is a *construction*, it is an increase of the organisation and cognitive complexity of the system. This process of emergent increment of order through the internal construction of functional organisations and simultaneous classification of the environment is a process of *self-organisation* (von Foerster, 1960, 1981).

### *2.2 Embodied constructions in self-organising systems*

Overall, it could be said that in the framework of second-order cybernetics, in contrary to the traditional cognitivist frameworks (Fodor, 1975, 1990; Newell, 1980), cognition is not considered as a process of gathering and assembling of representations that are directly related to objects or/and states of affairs of the environment. Instead, cognition is considered as a process of constant alteration of the intentional behavior of the system through the continuous modification of its functional organisation. In other words, a self-organising system is able to both establish and change its functionality in order to interact with an environment. This provides the self-organising cognitive system with a kind of autonomy that is not supported in the classical symbolic/cognitivist frameworks, since in the latter, any functional change would be externally imposed.

Furthermore, the nature of the systemic closure means that all the interactive alternatives of the cognitive system are internally generated and their selection is an entirely internal process. Therefore, such autonomous cognitive systems must construct their reality by using internally available structures. One should notice that the respective self-organised structures (eigenvalues) are specific to the particularities of the functionality of the cognitive system. Specifically, the functionality of the cognitive system is entirely dependent on its structural components and their interrelationships that establish the respective dynamics. Hence, the functionality of the cognitive system is immediately related to the maintenance of its systemic coherence (Collier, 1999), and consequently of its self-organisational dynamics (Collier and Muller, 1998).

This inclination of a self-organising cognitive system to maintain its own self-organisation constitutes the core of its intentional and purposeful (goal-oriented) interaction with the environment. This is a strong notion of embodiment based on the dynamics of the functional organisation of the cognitive system and it is quite different from the almost disembodied nature of a purely symbolic system.

### *2.3 Embodied constructions are not enough for the enhancement of autonomy*

However, this specific kind of embodiment and the consequent autonomy do not come gratuitously. The self-organising system cannot grasp every aspect of the environment

but only these aspects that can be constructed by its functional dynamics. Therefore, the meaning constructed by such an autonomous cognitive system is not open-ended. On the other hand, as it was previously noted, the design process is open-ended and emerges out of ill-defined goals and purposes of its participants, while it also results in ill-defined outcomes with ill-defined consequences.

This means that the anticipatory content of each self-organising system engaging in the design process should be open for revision and evolution, in order to reflect both this ill-definedness and the open-endedness as well. In this way, the self-organising system will have the ability to emerge new functions that will be directed towards new goals and hence, the new functionality will contribute to the autonomy of the system in new ways (Collier, 1999).

The need for open-endedness calls for the interaction of the self-organising system with other self-organising systems of the environment, while, the functional aspects of such an embodiment and its anticipatory content call for interactive emergent representations.

#### *2.4 From embodied constructions to interactive emergent representations*

*2.4.1 The need for interactive representations in autonomous cognitive systems engaging in a design process.* The internal constructions, with which the self-organising system classifies the environment and acts on it, are not representations of the environment. As von Glasersfeld (1995) argues, these constructions are *re-presentations* that are generated by the cognitive system in its embodied interaction with the environment. In second-order cybernetics, memory is understood as a process of re-presenting and re-membering by bringing past experiences into the present (von Foerster, 1969, 2003). Hence, *re-presentation* refers to the self-organised dynamics, by virtue of which a previous construction is re-constructed (re-presented) from memory, given that there is some sensory interaction (perturbation) with the environment.

In general, in the context of second-order cybernetics, the notion of representation as an encoded information, which is in an exact correspondence with the aspects of the environment that are supposed to be represented, is totally rejected. Actually, second-order cybernetic systems admit no functional usefulness to representations and they regard information only as socially ascribed to a process from other observers (Maturana and Varela, 1980; von Glasersfeld, 1995).

This rejection somewhat constrains the autonomy of self-organised systems to its internal dynamics. But, besides that, there are some cases, where the use of representations is demanded. Hence, Clark and Toribio (1994) argue in favor of “representational hungry” phenomena, which are mostly usual in the daily action of cognitive agents. In a more inclusive manner, Bickhard and Terveen (1995) are note some characteristic cases, where a kind of interactive representations that make possible the internal generation of error, which is detectable by the system itself, is necessary for the successful functioning of the cognitive system. In these cases, “the processing in the system must be potentially controllable, at least in part, by system error . . .” (Bickhard and Terveen, 1995, p. 210). Such cases appear, among others, in goal-directed interactions, “when system implicit anticipation of the courses and outcomes of interactions cannot be assured” (Bickhard and Terveen, 1995, p. 211) and in learning processes, as

Learning cannot be fully successfully anticipatory – if it were, there would be nothing to be learned. Learning must involve the possibility of error, and such error must be functionally detectable by the system itself so that the learning can be guided by it (Bickhard and Terveen, 1995).

Another case where interactive representations are needed is when there is more than one possible course of interaction for a specific environment and the system should choose among them on the basis of each anticipated outcomes of these interactions (Bickhard and Terveen, 1995; Bickhard, 2001).

It is apparent that the design process is more deeply implicated in the circumstances of a cognitive system-environment interaction, since a design outcome can be fulfilled in more than one way and its use (its interaction with one or more users, that is, other cognitive systems in the environment) may have more than one consequences. The selections cannot be realized through simple triggering, but some more complex process should be involved in the selections of the course of the interaction. Of course, there are some cases, where particular sensory interaction are known to provoke specific responses, especially in well-defined anticipation, where there is no need for the cognitive system to be aware of the subsequent internal outcomes. However, these are quite different and cannot provide an explanation for intentional and purposeful interaction of the autonomous cognitive system. Something is needed, that will justify the relation of internally self-organising structures of the autonomous cognitive system to particular aspects of its interaction with certain state of affairs in the environment.

*2.4.2 Emergent representations in an autonomous cognitive system.* Such “informational” internal states, which refer to certain conditions of the environment need to have an embodied and situated character (Moreno *et al.*, 1997), in order to be able to ground the representation to the context of the situated interaction between the autonomous cognitive system and the environment. Indeed, considering the functional closure of a self-organising system, its constructions can be seen as internal *in-formational* patterns, which have nothing to do with the transference of ontological information from the environment to the cognitive system. As long as this internal construction permits the cognitive system to survive, at least in this specific environment, and hence, to maintain or even enhance its autonomy, this construction should be considered as a *representation* of the situated interaction of the cognitive system with the respective environment.

Bickhard (1993, 2000, 2001) exemplifies this situation by postulating a recursive self-maintenant system, which is a self-organising system that has more than one means at its disposal in order to maintain its ability of being self-maintenant in various environmental conditions. This is a self-organising system which avoids going to equilibrium by continuously interacting with the environment, from where it finds the *appropriate conditions* for the success of its functional processes. Therefore, the primary goal of such a self-organising system is to maintain its autonomy in the course of interactions. Since, it is a self-organising system, its embodiment is of a kind that its functionality is immediately related to its autonomy, through the fact that its apparent inclination to maintain its autonomy, in terms of its self-maintenance (its purpose), constitutes the intentionality of its actions and hence, of its interaction with the environment.

In this way, the function of the cognitive system is guided by its autonomy, in the sense of the former contributing for the maintenance of the latter, while its

intentionality derives from this specific functionality, as the latter is being directed towards the primary purpose of maintaining the self-maintenance. What is still missing is meaning, on the basis of which the cognitive system decides which of the available functional processes should make use of, in order to successfully interact with a specific environment, that is, in order to fulfill its goal. But, where exactly is this meaning to be found?

Bickhard argues that such an autonomous system should have a way of differentiating the environments with which it interacts, and a switching mechanism in order to choose among the appropriate internal functional processes that it will use in the interaction. The differentiations are implicitly and interactively defined, as the internal outcomes of the interaction. These differentiations can occur in any interaction and the outcome of the interaction depends on the organisation of the participating subsystems and of the environment. Bickhard emphasizes that such differentiations create an epistemic contact with the environment, but they do not carry, in any way, any representational content, thus they are not representations by themselves. Rather, they are indications of the interactive potentiality of the functional processes of the autonomous cognitive system itself.

More specifically, the role of these differentiations is twofold:

- (1) they indicate the range of interactions that are functionally available for the cognitive system to use in this specific environment, that is, they indicate which further interactions might be possible or appropriate (Bickhard, 2000), in terms of at least contributing to the maintenance of the autonomy of the cognitive system; and
- (2) they implicitly predicate the environmental properties that would support the success of the functionally indicated interactive processes.

In other words, such differentiations functionally indicate that some type of interaction is available in the specific environment and hence, implicitly predicate that the environment exhibits the *appropriate conditions* for the success of the indicated interaction.

In this model, such differentiated indications constitute *emergent representations*. The conditions of the environment that are functionally and implicitly predicated by the differentiation, as well as, the internal conditions of the autonomous cognitive system (i.e. other functional processes or conditions), that are supposed to be supporting the selected type of interaction, constitute the *dynamic presuppositions* of the functional processes that will guide the interaction. These presuppositions constitute the *representational content* of the autonomous cognitive system regarding the differentiated environment. This content emerges in the interaction of the system with the environment and it corresponds to the *implicitly defined supports* of the functionally indicated interactive process (Bickhard, 2000).

This content may be in error, which means that the respective dynamic presuppositions may not hold (i.e. the environment may not provide the presupposed conditions). But this error will be functionally detectable by the autonomous cognitive system itself, since it will be functionally evaluated on the basis of the maintenance of the autonomy of the system (i.e. the indications of the content are embedded in the functionality of the system). These autonomous systems exhibit what Collier (1999, 2000) have called as *process* and *interaction closure*, namely a situation where the

internal outcomes of the interactions of the cognitive system with its environment contributes to the maintenance of the functional (constructive/interactive) processes of the system that are responsible for these specific interactions. Hence, meaning is produced by the functional evaluation of the representational content, internally in the autonomous cognitive system, but in the interaction of the system with its environment. It is in this way that meaning is a prerequisite and contributes to the maintenance of the autonomy of the cognitive system during its intentional and purposeful interaction.

In this perspective, each referential state of the autonomous cognitive system should be considered as situated in the context of the self-organised *in-formational* structures, as these are internally constructed due to its functional/organisational closure. In particular, these in-formational structures determine the intentional and purposeful interaction of the autonomous cognitive system, based on the variety of the indicated organisational forms they can support. Therefore, these in-formational structures indicate the representations that emerge (and hence, they can only be defined) in the context of the interaction of the autonomous cognitive system with the environment. In other words, any representational functional organisation is an emergent product of the interaction between the autonomous system and its environment. Hence, in an autonomous system, functionality provides intentionality simply because its functional structure carries, during the interaction, potentially reliable content about the environment.

The way an autonomous cognitive system uses its own functions in order to intentionally interact with the environment has some very interesting consequences regarding the design process, which are presented in the next section.

## *2.5 The design process as interaction between autonomous cognitive systems*

*2.5.1 Defining the design process and the design system.* Following on from the analysis made above, each autonomous cognitive system participating in the design process is considered as a self-organising system with the ability to maintain its autonomy in terms of its self-maintenance in different and dynamic environments. In a serial description (applicable only for demonstrative purposes) of the design process, each one of the participating autonomous systems could be defined as design-systems or user-systems at different time instances. However, the systemic and interactive approach adopted in this paper calls for a more participative and cooperative term, such as “user-designer” (called as UD, hereinafter), used by Banathy (1996, p. 226) to denote the “designing within the system” approach to design. Hence, an autonomous cognitive system acquires the identity of a UD system the very moment that it intentionally decides to engage in a design process. Consequently, in the framework described so far, *the design process is seen as an interaction between two or more autonomous UD systems, in order to maintain their capacity for self-maintenance, or in other words, in order to maintain the type of autonomy that permits them to internally create representational content.*

Since, in the analysis sketched before, autonomy guides functionality, the functional aspect of the design process in which each UD system interactively participates, becomes the purposeful and ongoing transformation and expansion of their already existing representations. For each UD system, a different representational content is internally emerging from their mutual attempts to incorporate the results of each other

actions (the artifact in each instance of the design process), as a perturbation and not as a static informational structure nor as a content in itself, into their functional organisation. Additionally, the group of such autonomous UD systems engaging in the design process constitutes a design system, which, as expected from the interactive nature of the design process, it is defined on the communicative/co-operative level[2].

*2.5.2 Describing the design process: the role of ill-defined-goals.* A logical sequence of the interaction cannot be implied, but for the benefit of this analysis, it can be said, that a UD system attempts to communicate its representations, regarding a possible solution towards an ill-defined goal, to the other UD systems participating in the design process, via the creation of an artifact. Considering the participative and co-operative aspects of the design process, the aim of this communication is to induce, in the other UD systems, the emergence of the necessary representational content that will guide their functional organisation towards the ill-defined goal.

From the perspective of autonomy, the aim of this communication, *from the point of view of the UD system that decides to communicate an artifact*, is to indirectly enhance the variety of the environment, so that the interaction of the UD system with this environment will facilitate the emergence of richer representational content that will further enhance its autonomy. Since, as discussed before, the representational content of each autonomous cognitive system partly depends on the dynamic presuppositions provided by the environment with which it chooses to interact, and partly on the functional dynamics of the system itself, the only way for an autonomous cognitive system to enhance its content is to provide for the enhancement of the representational content of all the other participants in the design process. Furthermore, this mutual enhancement should take place towards the direction of the specific ill-defined goal, since, according to our framework, its attainment will implicitly enhance the autonomy of the cognitive system.

Initially, in the early stages of an autonomous cognitive system such mutual dependence upon an ill-defined goal can be easily achieved. It becomes harder as long as ill-defined goals become more complicated. This happens when different cognitive systems construct different meanings of the design problem and provide different outcomes as possible solutions. This means that the ill-defined goal of the design process will never have a genuine and mutual recognition between its participants. Indeed, the degree of mutuality will decrease as far as the ill-defined goal becomes more complicated[3]. On this basis, it can be concluded that *the design process is the purposeful communication between two or more autonomous UD systems, in order to shape their dynamical interaction with the environment, in ways that they achieve a kind of functionality that contributes to the enhancement of their autonomy, by attempting to direct their functional organisations (i.e. themselves) towards an allegedly common ill-defined goal.*

At this point, it has been argued that two or more self-organising systems engage in an intentional and purposeful interaction with each other, in order to maintain and enhance their autonomy. In other words, self-organising systems engage in a design process out of necessity. From an observer's point of view, the design process could be considered as the attempt of two or more cognitive agents to provide each other a specific solution regarding a specific problem. In the interactive framework of second-order cybernetics, *the design process should be seen as an attempt of two or more autonomous systems to communicate their representational content regarding a*

*possible solution to an ill-defined goal – which is internally and differently formulated by each autonomous system – in order to maintain and enhance their autonomy*[4].

What needs to be clarified is the way this enhancement takes place in the face of complicated ill-defined goals and particularly, how the design process might acquire a greater directionality towards these goals, a case which will eventually contribute to the autonomy of the participants.

### **3. Anticipation and the design process**

#### *3.1 Anticipation and functionality*

Anticipation relates the present action of a cognitive system with its future state. An anticipatory system has the ability to organise its functional state, in such a way that its current behavior will provide the ability to successfully interact with its environment in the future. An anticipatory system needs to be able to take into consideration the possible results of its actions in advance (that is, prior to its action), hence, anticipation is immediately related to the meaning of the representations of the autonomous cognitive system (Collier, 1999). In this way, anticipation is one of the most characteristic aspects of autonomous systems due to their need to shape their dynamic interaction with the environment so as to achieve future outcomes (goals of the system) that will enhance their autonomy. In the context of the autonomous systems discussed so far, these future outcomes should satisfy the demand for process and interaction closure of the system.

Process and interaction closure are evaluated on the basis of the functional outcomes of the autonomous system, therefore, anticipation is immediately related to functionality (Collier, 2000). Even the simplest function requires anticipation in order to be effective. As mentioned before, anticipation is goal-directed. As a matter of fact, anticipation almost always requires functionality, which is, by default, a goal-oriented process. In this perspective, anticipation guides the functionality of the system through its representational content.

In the model of the emergence of representations in the special case of autonomous self-organising systems presented above, the representational content emerges in system's anticipation of interactive capabilities (Bickhard, 2001). In other words, the interactive capabilities are constituted as anticipation and it is this anticipation that could be inappropriate and this is detectable by the system itself, since such anticipation is embedded in the functional context of a goal-directed system. This type of anticipation is very different from the one supported by the cognitivist models of representation, which are trying to find a mapping of the environment to their past decisions. Here, the activity is future-oriented and it can be inappropriate, if the chosen interactive strategy does not internally yield the desired results, or if the respective environment does not support the type of interaction that would lead to the anticipated internal outcome.

#### *3.2 Dynamic anticipation directing the design process*

As stated before, the design process is open-ended and emerges out of ill-defined goals and purposes of its participants (the autonomous cognitive systems), while it also results in ill-defined outcomes with ill-defined consequences. The anticipatory content of each autonomous system engaging in the design process should be open for revision and evolution. Considering the dynamic and future-oriented type of anticipation

described above, it can be said that *each UD system participating in a design process should have the capability for anticipative interaction with the environment, in order to achieve the closure conditions that will contribute to its autonomy.*

As already said, the only way for an autonomous system to enhance its autonomy is by constructing even more adaptive representations towards its ill-defined goal. But this can only be achieved through the enhancement of its environment, that is, the emergence of new and more complex representations in the other UD systems, which belong into the same design system. If this is to happen towards the direction dictated by the otherwise subjectively formulated ill-defined goal, then the ability of each one UD system to anticipate the variety of the functional structures of all the other UD systems is crucial for the enhancement of the autonomy. Actually, the higher the degree of anticipation in each UD system, the higher its capacity to evaluate its interaction and the greater its ability to incorporate multiple possibilities in its performance, and also, the higher its capacity to consider the ill-defined consequences of the outcome of the design process, that is, the multiple ways in which each one of the other UD systems may choose to interact with the artifact.

In general, it can be said that the more the representational content of an autonomous system is evolved, the more dynamic its anticipative structures become (Collier, 1999; Bickhard, 2001). This has a positive effect in the anticipatory capacity of the autonomous system and in its capacity to evaluate its future interactions. The increase of the system's capacity for dynamic anticipation expands that what Christensen and Hooker (2000) call the *anticipatory time window*, which provides a certain degree of directionality (Christensen and Hooker, 2002) in the goal-directed interaction of the autonomous system. Overall, these capacities result in the emergence of new cognitive abilities for the autonomous system, thus, implicitly increasing its interactive autonomy.

Nevertheless, no matter how large the window of anticipatory interaction may be, all possibilities and selections regarding the outcomes and the ill-defined consequences of the design process cannot be inherent in the organisation of each UD system. A possible solution is that the UD system should evolve learning capabilities. This would provide the way to expand its dynamical anticipation capacity and its ability to evaluate a possible interaction. The UD system becomes less dependent and more sensitive regarding its contextual interactive capabilities. It increases its ability to better recognize its environment, evaluate conditions and better formulate its goal regarding the problem. This provides an infrastructure better suited to the UD system to define the design problem and anticipate the possibility of success in the emergent interactions between the other UD systems and the communicated artifact. Structural coupling is strengthened and the new and more adaptive representational content acquires a more prosperous field of emergence. Consequently, autonomy is increased.

However, it should be clear that not every external perturbation is useful for a dynamical anticipative interacting UD system. Only those contributing to the system's closure and therefore to the preservation of its autonomy would be selected for further exploitation. Since, in the proposed framework, closure is achieved at the level of differentiations and of the respective emergent representational content, it is concluded that autonomy cannot be statically identified, but as Collier (2000, 2002) suggests, it has a gradual nature.

Hence, *autonomy should be considered as an anticipative and future-directed property and it is a vital asset being directly related to the variety with which the UD systems participating in the design process will internally create adaptive emergent representations towards their ill-defined goals.* The artefacts are not objects any more, but interfaces functioning as triggers that drive the formation of new representational content. These interfaces between the UD systems should be seen as signals from one to the other that do not have a direct informational content in themselves. Rather, each UD system should exploit each artifact, as both a means of maintenance and the source of the enhancement of its own autonomy.

The consequence of this perspective is the paradigm shift from focusing on designing static things to focusing on designing the emergence of thoughts and of novel representational content. The interaction with an artifact results in a differentiated indication of the interactive capabilities of each UD system engaging in the design process. Taking this perspective, *autonomy depends on the degree to which the communicated representational content of each UD system, through the artifact, generates to the other UD systems the proper indications of the potentialities of their interactive capabilities.* In this way, the increase of autonomy is the result of a creative design process (Arnellos *et al.*, 2007).

What should be noted at this point is that based on this perspective, the content of the design process is not the artifact itself. It is also not static, since it is the attempt to communicate the UD system's representational content to the other UD systems actively participating in the design process. Moreover, due to the capacity for directed interaction, all UD systems engage in a mutual dependence with each other, while they are trying to increase their anticipatory capacity, no matter the degree of mutual recognition of their ill-defined goals. In their attempt to create richer representational structures towards their ill-defined goals, they are continuously interacting with the artefacts and hence, they learn to anticipate, or as it is suggested by Bickhard (2001) they anticipate the necessity to acquire new anticipations. Furthermore, the progressively increasing capability of the UD system's anticipation creates, as well, an intentional capacity. This is not the same as the traditional notion of intentionality considered as the sum of all system's representations. Intentionality derives from the UD system's functional capability of anticipative and purposeful interaction, and aims at the enhancement of each UD system's autonomy.

#### 4. Conclusions

The design process needs autonomous and interactive cognitive systems, while it is an ill-defined process with ill-defined and open-ended goals and consequences. It has been argued that the design process should be examined, analyzed and modeled in a framework of second-order cybernetics, as a cognitive system needs the respective embodiment in order to support the abovementioned characteristics and in order to be able to engage in a design process with its environment. In this case, the interaction of the cognitive system is guided by its self-organising functionality, which arises from its autonomy and it is directed towards the maintenance or/and the enhancement of this autonomy.

In a dynamic environment, an autonomous cognitive system with the ability to maintain its autonomy regarding its self-maintenance requires the internal generation of representational content that will drive its goal-oriented interactions.

The representational content emerges in the respective interactions and it depends upon the dynamic conditions of the environment and of the cognitive system itself. This content emerges in the form of anticipation that indicates the possibility of future interactions for the cognitive system and which result in the emergence of new functionality which in turn is directed towards new goals. The autonomous cognitive system will continue to interact with the environment towards these new goals, having as a primary aim the maintenance of its own autonomy.

The need for the enhancement of the autonomy makes each cognitive system engage in intentional and purposeful interactions with each other, aiming primarily at a common ill-defined goal. This is an interactive design process, which is conducted as a purposeful communication between two or more autonomous self-organising systems. These systems become user-designers and form a design system. Each user-design steps into this design system out of the necessity to maintain and enhance its autonomy.

Under the perspective of second-order cybernetics the design process is mutual, as in order for a user-designer system to be able to enhance its autonomy, it should first of all enhance the autonomy of its environment, that is, the autonomy of the other participants in the design process. This enhancement is goal-directed (hence, it is essentially a future-oriented process), but, each goal is differently and subjectively formulated in each autonomous cognitive system. This may provide some problems in terms of the directionality of the design process, that will immediately be reflected in the degree of enhancement of the autonomy of the respective cognitive systems.

These problems are smoothed out by the development of more elaborated anticipations which provide the autonomous system the capacity to evaluate its interaction and to anticipate its multiple possibilities. This provides a certain degree of directionality to each autonomous cognitive system participating in the design process, which brings each user-designer system closer to the ill-defined goal.

Finally, the capacity for directed interaction provides the capacity for learning, which prepares each user-designer to engage in more demanding and more complicated design processes. The prerequisites for learning is that the anticipatory content of the system should be open for revision, it should be able to be in error and this error should be internally detectable by the system itself. These properties are provided by a representational content that emerges in an autonomous system, which is cybernetically embodied, but it also has the ability to interact with the environment, in order to maintain its autonomy. Autonomy drives the design process and profits from it, when both the constructive and the interactive aspects of each participating cognitive system are considered.

#### Notes

1. See §2.5.1 for a clarification regarding the design process and the design system.
2. In the systems bibliography (Churchman, 1971; Ackoff, 1974, 1981; Banathy, 1989, 1996, 1998, 2000, etc.) the design process and consequently, the design system itself are directly defined at the social level. One can also talk about a cognitive system which comes forward to a design process and it is considered as a design system from the moment that it decides to engage in purposeful interactions with its environment (i.e. with other cognitive systems). However, from a systemic point of view, it seems to be more correct to consider as a design

system the set of all cognitive systems which are intentionally engaging in interactive design processes. In this view, and given that for the social scientists a cooperation is the co-action of two or more social actors, which is mediated by acts of communication, which in turn are mediated by acts of cognition by individual cognitive agents (Fuchs, 2003), the design system is solely defined at the social (cooperative) level and the design process acquires an interactive nature.

3. As a matter of fact, in the framework of second-order cybernetics there may be no goal at all (see for instance Glanville, 2004 for such a radical analysis). In the present paper, goals are considered as ill-defined and they are used in order to justify the directionality of the design process, through the intentionality of each autonomous cognitive system that belongs to the respective design system.
4. A reviewer has pointed out to us that the conclusion of this analysis is that design is living. We would like to point out that we fully agree with this remark, but the reason why this has not been stated that clearly so far is that we feel that a more detailed and in depth analysis is needed regarding the properties of the autonomous (living) systems that engage in a design process or the reader may be easily confused. However, we shall aim towards this direction in a future research, as we think that it is the most appropriate way for a naturalized account of the design process. The present paper aims at setting the basis for such an analysis and modelings.

## References

- Ackoff, R.L. (1974), *Redesigning the Future*, Wiley, New York, NY.
- Ackoff, R.L. (1981), *Creating the Corporate Future*, Wiley, New York, NY.
- Arnellos, A., Spyrou, T. and Darzentas, J. (2007), "Exploring creativity in the design process: a systems-semiotic perspective", *Cybernetics and Human Knowing*, Vol. 14 No. 1, pp. 37-64.
- Banathy, B.H. (1989), "The design of evolutionary guidance systems", *Systems Research*, Vol. 6 No. 4, pp. 289-95.
- Banathy, B.H. (1996), *Designing Social Systems in a Changing World*, Plenum, New York, NY.
- Banathy, B.H. (1998), "Evolution guided by design: a systems perspective", *Systems Research and Behavioral Science*, Vol. 15, pp. 161-72.
- Banathy, B.H. (2000), *Guided Societal Evolution: A Systems View*, Kluwer Academic/Plenum, New York, NY.
- Bickhard, M.H. (1993), "Representational content in humans and machines", *Journal of Experimental and Theoretical Artificial Intelligence*, Vol. 5, pp. 285-333.
- Bickhard, M.H. (2000), "Autonomy, function, and representation", *Communication and Cognition – Artificial Intelligence*, Vol. 17 Nos 3/4, pp. 111-31.
- Bickhard, M.H. (2001), "Function, anticipation, representation", in Dubois, D.M. (Ed.), *Computing Anticipatory Systems, CASYS 2000 – Fourth International Conference*, American Institute of Physics, Melville, NY, pp. 459-69.
- Bickhard, M.H. and Terveen, L. (1995), *Foundational Issues in Artificial Intelligence and Cognitive Science – Impasse and Solution*, Elsevier Scientific, Amsterdam.
- Christensen, W.D. and Hooker, C.A. (2000), "Anticipation in autonomous systems: foundations for a theory of embodied agents", *International Journal of Computing Anticipatory Systems*, Vol. 5, pp. 135-54.

- Christensen, W.D. and Hooker, C.A. (2002), "Self-directed agents", in MacIntosh, J. (Ed.), *Naturalism Evolution & Intentionality, Canadian Journal of Philosophy*, Vol. 27, pp. 19-52, Special Supplementary.
- Churchman, C.W. (1971), *The Design of Inquiring Systems: Basic Concepts of Systems and Organization*, Basic Books, New York, NY.
- Clark, A. and Toribio, J. (1994), "Doing without representing?", *Synthese*, Vol. 101, pp. 401-31.
- Collier, J. (1999), "Autonomy in anticipatory systems: significance for functionality, intentionality and meaning", in Dubois, D.M. (Ed.), *Computing Anticipatory Systems, CASYS'98 – Second International Conference, American Institute of Physics, Woodbury, New York, AIP Conference Proceedings*, Vol. 465, pp. 75-81.
- Collier, J. (2000), "Autonomy and process closure as the basis for functionality", in Chandler, J.L.R. and van de Vijver, G. (Eds), *Closure: Emergent Organizations and their Dynamics, of the Annals of the New York Academy of Science*, Vol. 901, pp. 280-91.
- Collier, J. (2002), "What is autonomy?", *International Journal of Computing Anticipatory Systems*, Vol. 12, pp. 212-21.
- Collier, J. and Muller, S. (1998), "The dynamical basis of emergence in natural hierarchies", in Farre, G. and Oksala, T. (Eds), *Emergence, Complexity, Hierarchy and Organization, Selected and Edited Papers from the ECHO III Conference, Acta Polytechnica Scandinavica, MA91 (Finish Academy of Technology)*.
- Fodor, J.A. (1975), *The Language of Thought*, Crowell, New York, NY.
- Fodor, J.A. (1990), *A Theory of Content*, MIT Press, Cambridge, MA.
- Friedman, K. (2003), "Theory construction in design research: criteria: approaches, and methods", *Design Studies*, Vol. 24 No. 6, pp. 507-22.
- Fuchs, C. (2003), "Co-operation and self-organisation", *tripleC*, Vol. 1 No. 1, pp. 1-52.
- Glanville, R. (2001), "A (cybernetic) musing: constructing my cybernetic world", *Cybernetics and Human Knowing*, Vol. 8 Nos 1/2, pp. 141-50.
- Glanville, R. (2004), "The purpose of second-order cybernetics", *Kybernetes*, Vol. 33 Nos 9/10, pp. 1379-86.
- Heylighen, A. and Bouwen, E.J. (1999), "Walking on a thin line – between passive knowledge and active knowing of components and concepts in architectural design", *Design Studies*, Vol. 20 No. 2, pp. 441-55.
- Jonas, W. (2001), "A scenario for design", *Design Issues*, Vol. 17 No. 2, pp. 64-80.
- Kampis, G. (2002), "The natural history of agents", in Gulyás, L., Tatai, G. and Váncza, J. (Eds), *Agents Everywhere*, Springer, Budapest, pp. 24-48.
- Laszlo, K.C. (2001), "Learning, design, and action: creating the conditions for evolutionary learning community", *Systems Research and Behavioral Science*, Vol. 18, pp. 379-91.
- Luhmann, N. (1995), "Why 'Systems Theory'", *Cybernetics & Human Knowing*, Vol. 3 No. 2, pp. 3-10.
- Maturana, H.R. and Varela, F.J. (1980), *Autopoiesis and Cognition: The Realization of the Living*, Reidel, Boston, MA.
- Moreno, A., Umerez, J. and Ibanez, J. (1997), "Cognition and life. The autonomy of cognition", *Brain & Cognition*, Vol. 34 No. 1, pp. 107-29.
- Nadin, M. (2000), "Anticipation: a spooky computation", *International Journal of Computing Anticipatory Systems*, Vol. 6, pp. 3-47.
- Newell, A. (1980), "Physical symbol systems", *Cognitive Science*, Vol. 4, pp. 135-83.

- Rocha, L.M. (1996), "Eigenbehavior and symbols", *Systems Research*, Vol. 13 No. 3, pp. 371-84.
- Ruiz-Mirazo, K. and Moreno, A. (2004), "Basic autonomy as a fundamental step in the synthesis of life", *Artificial Life*, Vol. 10, pp. 235-59.
- Simon, H.A. (1999), *The Sciences of the Artificial*, MIT Press, Cambridge, MA, (3rd rev. ed. 1996; Orig. ed. 1969; 2nd, rev. ed. 1981).
- von Foerster, H. (1960), "On self-organizing systems and their environments", pp. 1-19, reprinted in von Foerster H. (2003).
- von Foerster, H. (1969), "What is memory that it may hindsight and foresight as well?", in Bogoch, S. (Ed.), *Proceedings of the Third International Conference: The Future of the Brain Sciences*, New York, NY, pp. 19-64, reprinted in von Foerster, H., (2003), pp. 101-132.
- von Foerster, H. (1976), "Objects: tokens for (eigen-) behaviors", *ASC Cybernetics Forum*, Vol. 8, pp. 91-6, reprinted in: von Foerster, H., (2003), pp. 261-71 (Page numbers in the text refer to the reprint).
- von Foerster, H. (1981), *Observing Systems*, Intersystems Publications, Salinas, CA.
- von Foerster, H. (2003), *Understanding Understanding. Essays on Cybernetics and Cognition*, Springer-Verlag, New York, NY.
- von Glasersfeld, E. (1995), *Radical Constructivism: A Way of Knowing and Learning*, The Falmer Press, London.

#### About the authors

Argyris Arnellos holds a PhD from the University of the Aegean, Department of Product and Systems Design Engineering, a MSc in data communication systems from the Department of Electronic and Computer Engineering of Brunel University (UK) and a BSc in electronic engineering from the Department of Electronics of the Technological Educational Institution of Athens. His PhD thesis is entitled as "Exploring the Emergence of Meaning in Living Systems and in artificial environments" and it is an attempt to propose and establish a system-semiotic theoretical framework which is used for the modeling and analysis of the emergent meaning processes in the interactions at the biological, the cognitive and the social level. Since, 1998 he is a researcher in the Department of Product and Systems Design Engineering of the University of the Aegean, where he has been involved in a number of national and European research projects. He has published in scientific journals and participated in international and national conferences, in the areas of systems theory; 2nd-order cybernetics; semiotics; biosemiotics; design and design theory; information systems design; artificial intelligence; artificial life and human-computer interaction. Argyris Arnellos is the corresponding author and can be contacted at: arar@aegean.gr

Thomas Spyrou is an Assistant Professor in the Department of Product and Systems Design Engineering. His main interest is to research and to apply systems theories and approaches to real-world scenarios, especially in the case of information systems for human activity systems. He has over ten years of teaching experience in systems design, theories and methodologies of design, as well as various areas of human-computer interaction. He has been director or executive member of technical committees for the design of large-scale network and services infrastructures such as University of the Aegean network, Greek Secondary Educational network, Greek Universities network and Greek Research and Technology Network. He has served as part of the Ministry of the Aegean's Think Tank. He has directed and participated in a number of projects both funded nationally, and by the European Union. He has published in scientific journals and participated in conferences, in the areas of information systems design, holistic systems design, artificial intelligence, decision support systems, intelligent tutoring systems, simulation and security.

---

K  
36,9/10

**1224**

---

John Darzentas (BSc Athens Greece; MSc Sussex UK, PhD London UK) is Chair of Operational Research and Head of the Department of Product and Systems Design Engineering, University of the Aegean. He has held various academic positions in Britain, Finland and Greece, including lectureships at the Universities in London and Reading in the UK, visiting professorships at the University of Athens, and the Abo Akademi, in Turku, Finland. He has collaborated in and led many research projects, both in the UK and Greece as well as projects funded by the European Union on a range of subjects, including systems thinking; decision support; simulation; knowledge management; learning technologies human computer interaction; design; and lately design for all. He has spoken, and been invited to speak, at many conferences on various aspects of these topics, he is on the editorial board of several journals, and the author of a substantial number of papers in scientific journals and books.