# 17

# Naturalising the Design Process: Autonomy and Interaction as Core Features

ARGYRIS ARNELLOS, THOMAS SPYROU, JOHN DARZENTAS

## 1 Abstract/Introduction

This paper attempts to provide a naturalized description of the complex design process. The design process may be abstractly conceived as a future-creating activity that goes beyond 'facticity' and creates visions of a desirable future among groups of agents. It requires the engagement of individual or groups of cognitive systems in purposeful and intentional (meaning-based) interactions with their environment and consequently with each other. It is argued in this paper that a design process should be interactive, future-anticipatory and open-ended. Furthermore, a framework to explain and support the design process should have in turn its basis in a framework of cognition. It is suggested that the design process should primarily be examined within an interactive framework of agency based on 2nd order cybernetic epistemology. Future-oriented anticipation requires functionality which can be thought of as future-directed activity; indeed all but the simplest functionalities require anticipation in order to be effective. Based on the fundamental notions of closure, self-reference and self-organisation, a cybernetically-inspired systems-theoretic notion of autonomy is proposed. This conception of autonomy is immediately related to the anticipative functionality of the cognitive system, which constructs emergent representations while it interactively participates in a design process.

Consequently, the design process is seen as an interaction between two or more self-organising autonomous systems thereby constructing

*Beyond Description: Naturalism and Normativity.*
Marcin Miłkowski and Konrad Talmont-Kaminski (eds.).

ever more adaptive representations directed towards ill-defined outcomes. It is argued that this kind of autonomy is fundamental for the interactive establishment and definition of the design process as an essentially open-ended process.

## 2 Defining the Design Process

In the contemporary literature, there have been many efforts to define the design process and furthermore attempts to identify its most essential features. The nature and the purpose of the design process been much disputed and these issues have taxed various researchers over the last decades.

Defining the design process is certainly not an easy task and so, as one might expect, views are pretty diverse. As a matter of fact, Banathy (1996, pp. 11-13) lists up to twenty-four (24) definitions of design. Jones (1970) argues that design provides a means by which change is initiated in man-made things and in a somewhat parallel way Simon (1999, p. 111) states that, "…everyone designs who devises courses of action aimed at changing existing situations into preferred ones." 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. Glanville (2007, p. 1178) states that, "…design is an activity that is often carried out in the face of very complex (and conflicting) requirements…" In Glanville (2006) he furthermore agrees with Jonas (2007) in arguing that design can be considered the primary human activity.

All these definitions may differ from each other, but they appear to share a common opinion, namely, that the design process should be considered a cognitive activity. Furthermore, design should primarily be attributed to a cognitive agent and hence, it should have as its basis the cognitive process[1] (Arnellos, Spyrou, Darzentas, 2007a, 2007b).

But why attempt to give a definition of the design process? The main reason is that a proper definition will probably indicate the most suitable theoretical framework in which the design process can be thoroughly analyzed and explained. Thus far design activities have usually been considered a discourse situated between problem-framing and problem-solving (Simon, 1995). From this perspective, certain related activities include:

---

1 This is of course not limited only to humans.

the construction, elaboration and modification of the representations of the problem. In the problem-framing phase, designers refine their mental representations related to the problem, while during the problem-solving stage they elaborate their representations and then evaluate them (Bonnardel, 2000). Most often, the explanation of the activity of design is given by showing the ways an agent engaging in a design activity uses his representations (Simon, 1999) and the ways these representations are generated during the design activity. Regarding the latter, the main focus is on interaction with the environment and other design systems that play an important role in the generation of these representations (see e.g. Schön & Wiggins 1992, Gero & Kannengiesser 2004).

Is an explanation like this - based on the functionality of the problem-framing and problem-solving related representations - enough? That depends on what the main aims of such a framework are. At this point it should be noted that the rationale for an analytical framework to explain the design process is not to find a formalism to reduce the complexity of the design process, nor to produce models of structured representations to guide potential computer simulations. Such models would necessarily be much impoverished versions of reality, while any such framework would run into problems regarding contextuality and evolvability issues (Macmillan et. al., 2001) - it would be like trying to compute the transcomputable (Glanville, 2007).

However, as has been shown in (Arnellos, Spyrou, Darzentas, 2007a), an in-depth understanding of the complex and dynamic nature of a design process requires a framework to support the modelling of said processes so as to provide further understanding to provide better explanations and facilitate the emergence of creativity in the design process. Overall, it might be said that explaining the design process means trying to describe what the participating cognitive systems are doing and also, how they are doing whatever is that they do. In other words, a naturalized explanation of the design process should be provided.

## 3 What does it mean to naturalize the design process?

There are many different kinds of naturalism, but almost all of its adherents - especially those in favour of epistemological naturalism (see e.g. Feldman, 2006) – share the view that they provide different answers from those of traditional epistemologists to key epistemological questions such as the ways a cognitive system acts and the reasons for its actions.

Naturalization requires the justification of an explanation based on facts, i.e. based on natural relations or interactions. It is primarily an attempt to look inside the system under consideration and try to understand and explain how it works. This seems to be the most valid strategy for naturalism, as in this case the respective explanations can be objectively verified. Science is inherently progressive and so ‘scientific’ explanatory principles and rules should also be progressive; naturalization should have no end or a specific and discrete final state - it is an ongoing and open-ended process of scientific inquiry. Naturalization should be viewed as a constant process of reformulation of questions concerning a given phenomenon, making use of both quantitative and qualitative advances in the relevant field, thus aiming towards a better understanding, explanation, and description and modelling of this phenomenon.

We have argued elsewhere (Arnellos, Spyrou, Darzentas, 2007a, 2007b) that the design process has an interactive nature, realised at the social level by acts of communication, which in turn are mediated by acts of cognition by individual cognitive agents.[2] Hence, providing a naturalised description of the notion of agency (Arnellos, Spyrou, Darzentas, in press) could not be enough to amount to a naturalisation of the design process. Actually, a properly naturalised explanation of the design process should include the main characteristics of that process, characteristics which also have their source in the relevant communicative and social (cooperative) dimensions. These fundamental characteristics are briefly presented in the next section.

### *3.1 Ill-definedness and the Open-Ended Nature of the Design Process*

Most design problems are defined in terms of the features and needs of the people who will use the outcome of the design process (an artefact, which can be material or immaterial), the purpose it has for them and the form the artefact should posses in order to be deemed successful. Such design problems are ill-defined and the possible solutions are not clear

2 Another approach to the design process is the one which primarily considers the designer as a system which is able to engage in a circular and conversational process of creating innovative concepts and artefacts (e.g. Glanville, 2002, 2007). The approach adopted in this paper accounts for this case, as well as for the case where two or more cognitive agents are engaging in the design process. These two approaches are almost coinciding given that, during the design process, the respective agents are informed and reformed, irrespectively if one is designing on its own (i.e. by interacting with himself) or he is just in the phase of using (interacting with) the artefact of a design process.

from the beginning. Particularly, finding a solution also requires us to find out what the ‘real’ problem is, which in respect to human-centred problems is impossible. In such a complex process where solutions and problems co-evolve, one may well think that it is the solution that defines the problem. As such, the goal-oriented nature of the design process is usually related to a problem, or to a set of problems, the nature of which is constitutive of the design process itself.

On the other hand, there may be no ‘problem’ at all. Banathy (1996, p. 29) states that design confronts “...a system of problems rather than a collection of problems...” and he strongly argues in favour of viewing design as the attempt to find out what should become real, in terms of discerning what would be a desirable addition to the real word. From this perspective, the design process seems to be 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 always incomplete and imprecise, however much problem-solving advances. So, design problems are also open-ended. There are different logical paths leading to a design solution - different cognitive systems construct different meanings of the design problem and consequently provide different meaning-based outcomes as solutions. This makes designing into a process that is difficult to model and even more difficult to prescribe.

### *3.2 The Design Process Needs an Interactive Network*

The ill-defined and open-ended nature of a design problem makes both the goal and the respective constraints highly ambiguous. An internal evaluation of a possible solution is not enough – it would be subjective and disregard real-world needs. Internal evaluations of a closed system’s actions are bounded to its initial organizational complexity. The result will always be satisfactory for the system itself but rarely for its environment and hence for other systems. The lack of valuable information from the system at all stages of the design process is confronted by the opening of its boundaries to interact with the environment.

Banathy stresses the interactive and participatory nature of the design process arguing that a design system cannot design for others, but it can only design *with* others. Otherwise, as Lazlo (2001) also suggests, the design system cannot be said to engage in authentic design, it is merely trying to impose its visions and values, a situation represented by the conception of a design system operating solely in a cognitivist mode. As it is mentioned by Jonas (2001), there is a need to increase the internal complexity of a design system to deal with the increasing external complexity. Banathy (1996), Glanville (2001, 2007) and Arnellos, Spyrou, Darzentas (2007a, 2007b), all argue that the design process has a systemic nature implemented in an interactive and iterated mode. These two modes are both deemed necessary for they allow the testing of alternative solutions, the integration of insights, the formulation of viable strategies and attention to shifting parameters - factors that are all crucial to complex design processes.

### *3.3 The Content of the Design Process Is Not the Artefact Itself*

As is thoroughly analysed in (Arnellos, Spyrou, Darzentas, 2007a), the assignment of the design process to an interactive framework raises the question of the importance of the user of the design process outcome (the artefact). Users and stakeholders evaluate the artefact on the basis of their own individual experience. Considering that each user's experience and hence representational structures are different, the content of the design process should not be understood to be merely the artefact itself.

Indeed, the content should not be attributed to the aesthetic and practical properties of a fixed object (Kazmierczak, 2003). The content of the design process is subjectively interpreted and is changed by the user's cognitive processes while (s)he, in turn, is purposefully engaging in future design processes. The design system should now provide a form for dynamic and ill-defined content in such a way that will facilitate its creative interpretation by the user/receiver and, ultimately, other design systems.

From this perspective, validating the design outcome is an extremely difficult task, as there can be quite unexpected benefits for the other design systems issuing from the use of that outcome. So the next difficult question might be where the design process ends. Of course, as mentioned before, design is an open-ended and iterated process, and as such it never comes to an end. But the designer should be able to take an open-ended decision in this process and since the content under discussion is in fact a meaning understood by the participants; it would seem that their anticipation must have a key role in the design process.

### *3.4 The Design Process Is Future-Anticipative*

The varied interpretation of content from multiple receivers 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 has an anticipatory nature by which it will be placed in a pragmatic context and simultaneously be projected towards 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, a cognitive system which learns from the past and appraises what is presently useful and desirable by simultaneously projecting content into the future.

According to the fundamental characteristics of the design process presented above, its naturalized explanation should incorporate in its description the engagement of individual cognitive systems (agents) in intentional and purposeful interactions with their environment and consequently with each other – actions undertaken in order to be able to fulfil their ill-defined goals. As argued in (Arnellos, Spyrou, Darzentas, 2007b) cognitive systems like these should have an autonomy that will guide them through this kind of interaction, based on their open-ended anticipative functionality. As expected, a complete analysis and description of the design process in a naturalised framework is not an easy task. We therefore begin this endeavour with a naturalised account of agency presented in detail in (Arnellos, Spyrou, Darzentas, in press) and we will try to show how the main characteristics of the design process are constructively supported within the resulting framework. Autonomy and interactivity will turn out to be core features in this naturalised framework.

## 4 A Naturalized Description of Agency

### *4.1 Autonomy drives agency via the intentional creation of functional meaning*

There are many frameworks explaining agency in contemporary scientific literature and thus many diverse definitions of a cognitive agent. On the other hand, and as we have repeatedly suggested in (Arnellos, Spyrou, Darzentas, 2007b, 2008, in press) a strong notion of agency calls for: *interactivity* – 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 on its own resources.

This definition mentions three fundamental capacities that an agent should exhibit in a somewhat integrated way regarding their existence and their evolutionary development. According to this definition, agency requires interactivity – which in turn implies action on the environment. This action is not accidental but intentional; it is purposeful action directed towards a goal and it is driven by content such as beliefs and desires. In addition, an agent like this exhibits the property of autonomy as it interacts with the environment in an intentional manner based on its *own* resources, including its own internal content. These three properties seem quite interdependent, what becomes clear when one attempts to consider if it is possible for one of them to increase qualitatively while the others remain at the same level.

This interdependency is emphasized by Collier (1999) when he suggests that 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.

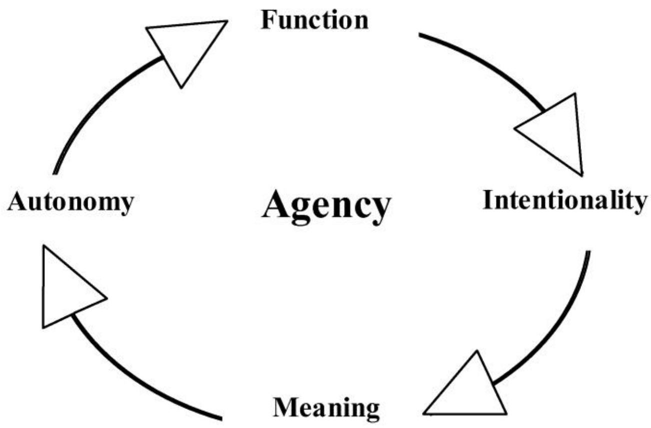


Figure 1. Interdependence of autonomy, functionality, intentionality and meaning in an autonomous agent.

These properties and their interdependence are characteristic of the strong notion of agency (i.e. they are features of living systems) – that which is emergent in the functional organisation of a 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 (see e.g. Ruiz-Mirazo and Moreno, 2004). In this sense, meaning should be linked with the functional structures of the system. So 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. As such, the enhancement of autonomy means setting certain goals by the system itself. So the intentionality of the system is guiding its behaviour through meaning.

It should be noted that in such a theoretical set-up, meaning and its functional substratum are the defining properties of an autonomous agent that can act intentionally. In other words, an autonomous system may act intentionally if its actions are mediated by meaning. The foundations of functional emergence of this sort have been established in the systems-theoretic framework of second-order cybernetics.

### *4.2 Closure and Self-Reference for Self-Organisation*

In second-order cybernetic epistemology, a cognitive system is able to carry out the fundamental actions of distinction and observation. It observes its boundaries and it is thus itself 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, with the ability to create new distinctions (actions) based on previous ones, to judge distinctions, and to increase its complexity by producing new meanings in order to interact. The self-referential loop can only exist in relation to an environment, but it nevertheless goes beyond 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* and *operational/organisational closure* (von Foerster, 1960, 1981).

Due to this closure, the self-reference of an observation produces meaning inside the cognitive system, which is used as a model for further observations in order to compensate for external complexity. Indeed, this closure is functional in so far as the effects produced by the cognitive system produce the maintenance of its systemic equilibrium through the emergence of more complex organizations. With system closure, environmental complexity is based solely on system observations, so 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; they are instead the results of recurrent cognitive functions in the structural coupling between the cognitive system and the environment. So each emergent function based on observations is a construction; it is an increase of the organisation and cognitive complexity of the agent. 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-organization* (von Foerster, 1960, 1981).

The nature of this systemic closure means that all the interactive alternatives of the cognitive system are internally generated and their selection is an entirely internal process. So, autonomous cognitive systems like this must construct their reality by using internally available structures. We should note that the respective self-organised structures (eigenvalues) are specific to the particular functionality of the cognitive system. Specifically, the functionality of the cognitive system is entirely dependent on its structural components and their interrelationships – interrelationships that establish the respective dynamics. So the functionality of the cognitive system is immediately related to the maintenance of its systemic cohesion (Collier, 1999), and consequently its self-organising dynamics. 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 pattern of self-organising dynamics requires a certain type of cohesion.

### *4.3 Cohesion via Process Closure for Self-Organisation*

Cohesion is an inclusive capacity of an autonomous system and it indicates the existence of causal interactions among the components of the system in which certain capacities emerge - the respective components are constituents for the system itself. As such, cohesion can only be explained with respect to the causal roles that the constituent components and the relations among them, that is, their functional processes, acquire in the dynamic organization of the system.

Cohesive systems exhibit different kinds of correlations between different processes with respect to the degree (or the type) of cohesion that they exhibit. Systems with very strong and highly local bonds exhibit powerful cohesion which does not necessarily provide them with genuine autonomy and agency. The essential type of cohesion emerges in systems that are thermodynamically open and function in conditions that are far from equilibrium (Collier and Hooker, 1999). As Collier (2007) has stressed, since there is an internal need in these systems for the coordination of the processes to be able to achieve viability (self-maintenance), one should expect to find holistic organization in which organizationally/operationally open aspects of lower level are closed at higher organizational levels. This is a highly constructive type of autonomy (see Arnellos, Spyrou, Darzentas, 2008) and it requires what Collier (1999) suggests we call *process closure* (in the convention of 'organizational/operational closure'), to stress the fact that in such autonomous systems there are some internal constraints controlling the internal flow of matter and energy, and by doing so, the whole system acquires the capacity to carry out the respective processes, since these processes will contribute to its self-maintenance.

The preceding analysis of agency is a useful basis for understanding and explaining the design process. However, the description given is not adequately naturalized. Agency comes in degrees and at various levels in nature (Arnellos, Spyrou, Darzentas, in press). Genuine autonomous systems, such as living systems, are distinguished by a high degree of disentanglement from the environment, not in terms of their interactive processes, but on the contrary, in terms of their ability to adapt in various environmental disturbances. The systems described so far seem to exhibit a functional organization that is too tightly connected with their environments, but with minimal interactive characteristics, and as such, they cannot evolve beyond a certain threshold. Such systems are at the threshold of achieving autonomy, but at most this is a reactive kind of agency, insufficient to be able to engage in genuine design processes, where interaction is a vital asset.

Agency and the design process cannot be solely a matter of internal constructive processes and process closure. The need for open-endedness calls for interaction of the autonomous agent with the environment, while, the functional aspects of such an embodiment and its anticipatory content calls for advanced and efficient mechanisms of controlling and purposefully managing these interactions.

### *4.4 Interaction Closure and the Emergence of Normative Functionality*

In section 4.2 it was mentioned that any system (defined and considered under the more general framework of second-order cybernetics) makes a distinction between the components that constitute itself and the rest of the elements that form its environment. The respective qualitative and quantitative imbalance indicates an asymmetry between a system and its environment. In the self-organizing systems described so far, this asymmetry is created and maintained by the functionality of the system through the establishment of internal constructive relations that organizationally differentiate the system from its environment and specify its autonomy and its identity.

This is an interpretive asymmetry (Hoffmeyer, 1998; Arnellos, Spyrou, Darzentas, in press) with some very interesting implications. Bickhard (1993, 2000) exemplifies the implications of this asymmetry by postulating a recursive self-maintenant system: a self-organizing system that has more than one means at its disposal to maintain its ability to be self-maintenant in various environmental conditions. This is a self-organizing system which functions far from thermodynamic equilibrium by continuously interacting with the environment, from where it finds the *appropriate conditions* for the success of its functional processes. Processes which are far from equilibrium cannot be kept in isolation, as they will lose their dynamic-functional stability.

So, the interactive opening of the system to the environment is considered the most important point in its evolution towards genuine autonomy and agency, as it first of all enhances the stability of the system and its ability to maintain its maintenance. Specifically, the interactions in which an autonomous agent engages will be *functional* and *dysfunctional* (Moreno and Barandiaran, 2004). The former corresponds to those interactions which are integrated in the functional organization of the agent and in this way they contribute to its self-maintenance. The latter corresponds to interactions that cannot be properly integrated in the functional organization and hence do not contribute or/nor disturb the self-maintenance of the system.

So the primary goal of such a self-organizing system is to maintain its autonomy in the course of its interactions. Since it is a self-organizing 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 its interaction with the environment. As such, autonomous systems do not only exhibit process closure, but also *interaction closure* (Collier, 1999, 2000, 2007), a situation where the internal outcomes of the interactions of the autonomous system with its environment contributes to the maintenance of the functional (constructive/interactive) processes of the system that are responsible for these specific interactions. It is cohesion via process and interaction closure that truly distinguishes autonomous systems from other kinds of cohesive systems. An autonomous system is not only able to maintain itself, but it can also meaningfully alter its internal functionality in order to adapt to complex and changing conditions around the environment.

This capacity for meaningful critique regarding the functional (the ‘good’) and the dysfunctional (the ‘bad’) with respect to the maintenance of the system is a normative one. Self-maintenant systems that exhibit *normative functionality* are truly autonomous systems and they present genuine agency. In this way, the overall functional closure (process and interaction closure) of an agent is guided by its autonomy - in the sense that the former contributes to the maintenance of the latter - while its intentionality derives from this specific normative functionality, as the latter is being directed towards the primary purpose of maintaining self-maintenance. This cohesive combination of process and interaction closure is responsible for the emergence of functional norms within the autonomous system and for the autonomous system itself.

These functional norms, in a way, attribute values of truth or falsity, and they are emergent in the system’s interactions with the environment. Particularly, they are internal constructions that attribute binary values of to processes and/or the interactions of an autonomous system. The binary nature does not imply explicitness, but on the contrary, it can be said that the higher the autonomy and agency of the cognitive system, the higher the degree of abstraction of the concepts to which some of its norms can be related. This means that even though norms are constructed by the autonomous system itself, there are too many cases where their interactive satisfaction is not immediately recognizable by the autonomous system.

It would appear, according to the strong notion of agency introduced in 4.1, we are still lacking *meaning* - on the basis of which the cognitive system decides which of the available functional processes to make use of in order to successfully interact with a specific environment, i.e. to fulfil its goal or to satisfy its functional norms. An autonomous system anticipates via the respective representational content (meaning). But where exactly is this content to be found? The naturalistic requirement for an explanation of the constructive and interactive aspects of normative functionality - i.e. of the efficient control and management of the constructive/interactive capabilities of an autonomous agent - calls for the introduction of interactive, emergent representations.

### *4.5 Anticipations and the Emergence of Representational Content*

Bickhard argues that an autonomous system of the aforementioned kind should have a way of differentiating the environments with which it interacts, and should possess a switching mechanism to choose among the appropriate internal functional processes that it will use in the interaction in question. The differentiations are implicitly and interactively defined, as the internal outcomes of the interaction - which in turn depend on the functional organization of the participating subsystems and on that of the environment. These differentiations create an epistemic contact with the environment, but they do not carry any representational content at all. However, they are indications of the interactive potentiality of the functional processes of the autonomous system itself. As such, these differentiations functionally indicate that some type of interaction is available in the specific environment and hence implicitly anticipate that the environment exhibits *appropriate conditions* for the success of the indicated interaction.

In this model (Bickhard, 1993; 2000), differentiated indications like the aforementioned constitute *emergent representations*. The conditions of the environment that are functionally and implicitly responded to 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. What remains to be shown is how this representational content is related to the anticipations of an autonomous system.

Anticipation relates the present action of an agent with its future state. An anticipatory system has the ability to organise its functional state, in such a way that its current behaviour will provide the ability to successfully interact with its environment in the future. Such a system needs to be able to take into consideration the possible results of its actions in advance, 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 and in general, for system's normative functionality.

Normative functionality is evaluated on the basis of the functional outcomes of the autonomous system; therefore, anticipation is immediately related to functionality (Collier, 2007). 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. From 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 an autonomous agent presented above, *the representational content emerges in a system's anticipation of interactive capabilities* (Bickhard, 2001). In other words, the interactive capabilities of a system are the subject of anticipation by the self-same system. This anticipation may be erroneous, potential errors being detectable by the system. Anticipation is an integral part of the functional context of a goal-directed system (a system which exhibits what we call 'emergent normativity').

These anticipations guide the interpretive interactions of an autonomous agent. In case these interactions contribute to the agent's self-maintenance, its capability for interactive anticipation progressively increases and as such its intentional capacity increases too (Christensen and Hooker, 2002; Arnellos, Spyrou, Darzentas, in press).

## 5 The Design Process as Interaction between Autonomous Systems

Following on from the analysis made above, each autonomous cognitive system participating in the design process is considered a self-organising system with the ability to maintain its autonomy in terms of its self-maintenance in different and dynamic environments. Hence, an autonomous cognitive system acquires the identity of a UD system[3] the very

[3] 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

moment that it intentionally decides to engage in a design process. Consequently, in the framework described so far, *the design process is viewed 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.*

In the analysis sketched before, autonomy guides functionality; so *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 artefact 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. In addition, *a group of autonomous UD systems such as these, 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*.

A logical sequence of the interaction cannot be implied, but for the benefit of this analysis let's just assume 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 artefact. 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 artefact*, is to indirectly enhance the degree of variety found in 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 in 4.5, the representational content of each autonomous cognitive system partly depends on the dynamic presuppositions provided by the environment with which it

paper calls for a more participative and cooperative term, such as 'user-designer' (called as UD, hereinafter), used by Banathy to denote the 'designing-within-the-system' approach to design (Banathy, 1996, p. 226) 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 in a way that furthers the achievement of the specific but ill-defined goal standing before the system, since, according to the framework we have adopted; its attainment will implicitly enhance the autonomy of the cognitive system.

### *5.1 The Role of Ill-Definedness*

Initially, in the early stages of an autonomous cognitive system this mutual dependence upon an ill-defined goal can be easily overcome. The achievement of goals becomes harder as those 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. On this basis, it can be concluded that *the design process is purposeful communication between two or more autonomous UD systems, in order to shape their dynamical interaction with the environment, so as to 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*.

The ways this enhancement takes place in the face of complicated ill-defined goals and the ways the design process might acquire a greater directionality towards these goals are discussed in the following section.

### *5.2 The Design Process Is Directed By Dynamic Anticipation of the Participating Autonomous Systems*

As mentioned above, the design process is open-ended and emerges out of the ill-defined goals and purposes of its participants (autonomous 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 to revision and evolution. Considering the dynamic and future-oriented type of anticipation described in 4.5, 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 in the same overall design system. If this is to move in the direction of 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 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 artefact.

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.

### *5.3 The Design Process as Learning*

Nevertheless, no matter how large the window of anticipatory interaction may be, not all possibilities and selections regarding the outcomes and the ill-defined consequences of the design process can be inherent in the organisation of each UD system. A possible solution to this predicament is for the UD system to evolve learning capabilities. This would provide a 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 for the UD system to be able to define the design problem and anticipate the possibility of success in the emergent interactions between the other UD systems and the communicated artefact. Structural coupling is strengthened and the new and more adaptive representational content acquires a more prosperous field of emergence. Consequently, autonomy is increased.

However, in the proposed framework closure is achieved at the level of differentiations and of the respective emergent representational content, so autonomy cannot be statically identified. Instead, as Collier (2000, 2002) suggests, it has a gradual nature. So *autonomy should be considered an anticipative and future-directed property and it is a vital asset 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. Therefore, each UD system should exploit each artefact, as both a means of maintenance and a source of enhancement of its own autonomy. A consequence of this point of view is a paradigm shift: from focusing on designing static things to focusing on designing the emergence of thoughts and of novel representational content. Interaction with an artefact results in a differentiated indication of the interactive capabilities of each UD system engaging in the design process. From this perspective, *autonomy depends on the degree to which the communicated representational content of each UD system, through* *the artefact, gives the other UD systems the proper indications of the potentialities of their interactive capabilities*. Ultimately, the increase of autonomy is the result of a creative design process (Arnellos, Spyrou and Darzentas, 2007a).

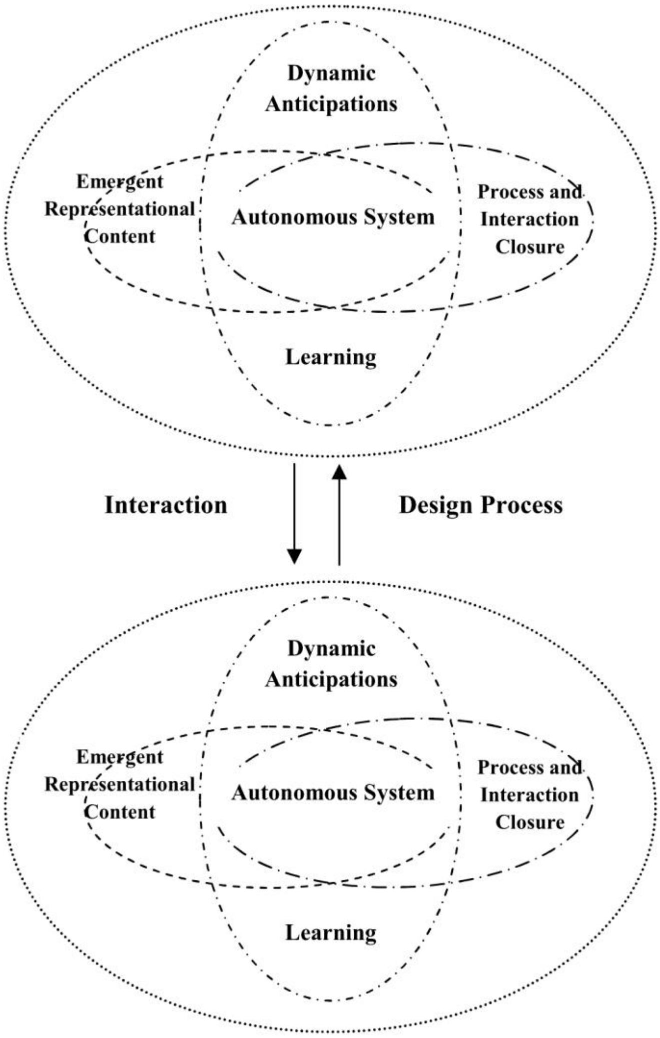


Figure 2. The design process as the interaction between two autonomous systems that guide their learning through the use of dynamic anticipations with emergent representational content.

The design process together with its key aspects, supported by its participating autonomous systems, is abstractly depicted in Fig. 2. What should be noted at this point is that from this perspective, the content of the design process is not the artefact itself. It is 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 also creates 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 to enhance each UD system's autonomy.

## 6 Conclusions

Naturalization is quite controversial: one has to proceed with a continuous formulation of questions about the phenomenon in question, taking into consideration both the quantitative and the qualitative progress of science with respect to relevant notions and beliefs. Design should have a cognitive foundation. Agency appears to be one of the most complicated capacities that nature presents and the quest for its naturalized explanation is not easy (Arnellos, Spyrou, Darzentas, in press).

An attempt has been made to provide a naturalized analysis of the design process by considering the latter as the purposeful interaction between two or more autonomous agents. The interaction of each cognitive system is guided by its self-organising functionality, which arises from its autonomy and it is directed towards the maintenance and/or the enhancement of this autonomy. The respective analysis departs from the systemic framework of second-order cybernetics, where an agent is viewed as a self-organising system exhibiting self-reference and organisation closure. The respective functional circularity is the result of interaction closure, which combined with process closure provides the means for the emergence of functional norms in each autonomous system.

This normative nature is grounded in values produced within the system and used in order to guide its interactions. As an agent evolves, some of its norms cannot be immediately identified and hence satisfied in its functional organization – so some mediation of their uncertain interactive potentialities is required. This mediation takes place through the formation of relevant anticipations with the respective representational content. These dynamic anticipations guide the autonomous system in the design process and provide the system with the ability to bring itself closer to its ill-defined goals. In this way, autonomy drives the interactive design process and at the same time profits from it.

## 7 References

Arnellos, Argyris, Spyrou, Thomas, Darzentas, John (2007a). "Exploring creativity in the design process: A systems-semiotic perspective". *Cybernetics and Human Knowing* 14 (1), pp. 37-64.

Arnellos, Argyris, Spyrou, Thomas, Darzentas, John (2007b). "Cybernetic embodiment and the role of autonomy in the design process". *Kybernetes* 36 (9/10), pp. 1207-1224.

Arnellos, Argyris, Spyrou, Thomas, Darzentas, John (in press) "Towards the Naturalization of Agency based on an Interactivist Account of Autonomy". *New Ideas in Psychology* (*doi:10.1016/j.newideapsych.2009.09.005*).

Arnellos, Argyris Spyrou, Thomas, Darzentas, John (2008). "Emergence and Downward Causation in Contemporary Artificial Agents: Implications for their Autonomy and Some Design Guidelines". *Cybernetics and Human Knowing* 15 (3-4), pp. 15-41.

Banathy, Bela H. (1996). *Designing Social Systems in a Changing World*, Plenum, New York.

Bickhard, Mark H. (1993). "Representational Content in Humans and Machines". *Journal of Experimental and Theoretical Artificial Intelligence* 5, pp. 285-333.

Bickhard, Mark H. (2000). "Autonomy, Function, and Representation". *Communication and Cognition — Artificial Intelligence* 17 (3-4), pp. 111-131.

Bickhard, Mark H. (2001). "Function, Anticipation, Representation". In *Computing Anticipatory Systems, CASYS 2000 - Fourth International Conference* ed. by D. M. Dubois, NY American Institute of Physics, Melville, pp. 459-469.

Bonnardel, Nathalie (2000). "Towards understanding and supporting creativity in design: analogies in a constrained cognitive environment". *Knowledge-Based Systems* 13, pp. 505–513.

Christensen, Wayne D. and Hooker, C. A. (2002). "Self-directed agents". *Canadian Journal of Philosophy*, Special Supplementary 27, pp. 19-52.

Collier, John (1999). “Autonomy in anticipatory systems: significance for functionality, intentionality and meaning”. In *Computing Anticipatory Systems, CASYS'98 - Second International Conference, American Institute of Physics* ed. by D. M. Dubois, New York: Woodbury, pp. 75-81.

Collier, John (2000). “Autonomy and Process Closure as the Basis for Functionality”. *Annals of the New York Academy of Science* 901, pp. 280-291.

Collier, John (2002). “What is Autonomy?”. *International Journal of Computing Anticipatory Systems* 12, pp. 212-221.

Collier, John (2008). “Simulating autonomous anticipation: The importance of Dubois’ conjecture”. *BioSystems* 91, 2, pp. 346-354.

Collier, John and Hooker C.A. (1999). “Complexly Organised Dynamical Systems”. *Open Systems and Information Dynamics* 6, pp. 241-302.

Feldman, Richard (2006). “Naturalized Epistemology”. In *The Stanford Encyclopaedia of Philosophy (Fall 2006 Edition)*, ed. by Edward N. Zalta, URL = http://plato.stanford.edu/archives/fall2006/entries/epistemology-naturalized/.

Friedman, Ken (2003). “Theory construction in design research: criteria: approaches, and methods”. *Design Studies* 24 (6), pp. 507-522.

Gero John S. & Kannengiesser, Udo (2004). “The situated function–behaviour–structure framework”. *Design Studies*, *25* (4), pp. 373-391.

Glanville, Ranulph (2001). “A (Cybernetic) Musing: Constructing my Cybernetic World”. *Cybernetics and Human Knowing* 8 (1-2), pp. 141-150.

Glanville, Ranulph (2006b). “Design and mentation: Piaget’s constant objects”. *The Radical Designist*, zero issue, available at: www.iade.pt/designist/jornal/jornal.html (accessed 14 April 2007).

Glanville, Ranulph (2007). “Try again. Fail again. Fail better: the cybernetics in design and the design in cybernetics”. *Kybernetes* 36 (9/10), pp. 1173-1206.

Hoffmeyer, Jesper (1998). “Surfaces inside surfaces. On the origin of agency and life”. *Cybernetics and Human Knowing* 5 (1), pp. 33–42.

Jonas, Wolfgang (2001). “A Scenario for Design”. *Design Issues* 17 (2), pp. 64-80.

Jonas, Wolfgang (2007). “Research through DESIGN through research. A cybernetic model of designing design foundations”. *Kybernetes* 36 (9/10), pp. 1362-1380.

Jones, John C. (1970). *Design methods: seeds of human futures*. London, New York: Wiley-Interscience.

Kazmierczak, Elżbieta T. (2003). “Design as Meaning Making: From Making Things to the Design of Thinking”. *Design Issues* 19 (2), pp. 45-59.

Macmillan, Sebastian, Steele, John, Austin, Simon, Kirby, Paul & Spence, Robin (2001). “Development and verification of a generic framework for conceptual design”. *Design Studies* 22 (2), pp. 169-191.

Moreno, Alvaro and Barandiaran, Xabier (2004). “A naturalized account of the inside-outside dichotomy”. *Philosophica* 73, pp. 11-26.

Nadin, Mihai (2000). “Anticipation: A Spooky Computation”. *International Journal of Computing Anticipatory Systems* 6, pp. 3-47.

Schön, Donald A. & Wiggins, Glenn (1992). “Kinds of seeing and their functions in designing”. *Design Studies* 13 (2), pp. 135-156.

Simon Herbert A. (1995). “Problem forming, problem finding and problem solving in design”. In *Design & Systems*, ed. by Arne Collen & Wojciech Gasparski. New Brunswick: Transaction Publishers, pp. 245–257.

Simon, Herbert A. (1999). *The sciences of the artificial*, The MIT Press, Cambridge, MA.

von Foerster, Heinz (1960). “On self-organizing systems and their environments”. Reprinted in von Foerster, Heinz (2003), pp. 1-19.

von Foerster, Heinz (1976). “Objects: tokens for (eigen-) behaviors”, *ASC Cybernetics Forum* 8, pp. 91-6. Reprinted in: von Foerster, Heinz (2003), pp. 261-71.

von Foerster, Heinz (1981). *Observing Systems*. CA: USA, Intersystems Publications.

von Foerster, Heinz (2003). *Understanding Understanding. Essays on Cybernetics and Cognition*. New York: Springer-Verlag.