Modeling as Autonomous System: Varela Meets Darwin

Paper to be presented at the
30th International System Dynamics Conference 2012,
St. Gallen, Switzerland

Martin Schaffernicht,
Facultad de Ciencias Empresariales
Universidad de Talca
Avenida Lircay s/n
3460000 Talca, CHILE
martin@utalca.cl

Camilo Olaya
Universidad Los Andes
Bogota, Colombia
colaya@uniandes.edu.co

Modeling as Autonomous System: Varela Meets Darwin

Paper to be presented at the

30th International System Dynamics Conference 2012,

St. Gallen, Switzerland

Abstract

This paper posits mental modeling as autopoietic systems of ideas based on processes of selection. A thought experiment is proposed to illustrate the argumentation. After reviewing the essential features of selection theory and mental models, we present modeling as driven by a process of variations and selections. We then show how the deliberate process of model development fosters the generation of mental variety and transfers the locus of selection from external to internal. We deduce that modeling, as far as it succeeds in framing successful actions, is a selective advantage; however, as such it may only play out over generations. If mental models are autonomous and enactive systems, the use of selection theory to describe their evolution is fruitful. We believe both theoretic bodies to be useful for further investigating how model develop in minds and how internal selectors can be designed such as to reduce the need for external selection.

Keywords: mental models, autonomy, Varela, autopoiesis, self-replication, selection theory.

1. INTRODUCTION

There are those who look at things the way they are, and ask why...
I dream of things that never were, and ask why not?
Robert Francis Kennedy

Maturana and Varela (1980:78; 1994:69) defined an autopoietic machine as "organized as a network of processes of production (transformation and destruction) of components which:

(i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and

(ii) constitute it (the machine) as a concrete unity in space in which they (the components) exist by specifying the topological domain of its realization as such a network. "

Two additional characteristics are (Maturana and Varela, 1994: 14) that as part of their operation, they

(iii) generate a border defining the network such as to operate as a logically closed unity separated from its environment and

(iv) configure a flow of molecules which, upon entry into the network become components of the network, and upon ceasing their participation in the network's operation, become part of the environment again.

Both authors have repeatedly expressed their skepticism with regards to the description of other, non-biological, phenomena as autopoietic systems (Maturana, 1980, 1988; Varela, 1979, 1981a). In (Maturana, Varela, 1980, p. 89) they argue that "[...] the space defined by an autopoietic system is self-contained and cannot be described by using dimensions that define another space. When we refer to our interactions with a concrete autopoietic system, however, we project this system on the space of our manipulations and make a description of this projection."

Strictly speaking, a cell is an autopoetic system since it generates its own material realization, which is hardly transferable to social systems which are third order entities (consisting of multi-cellular organisms consisting of cells). However, the defining characteristics of operational closure and structural coupling do apply to higher order systems. Indeed, "by characterizing multiple-order autopoietic systems without requiring physical definition of autopoietic systems, Maturana leaves open the possibility of defining operational systems (e.g. social) as autopoietic systems in their own right" (Fleischaker,1988, p. 41). Varela (1981b) has gone on to develop his view concerning the *autonomy* of living systems and the circular relationship of *enaction* between such systems and their milieu.

We believe that *modeling* – the deliberate development and use of mental models in action by the way of developing explicit models - qualifies as *autopoietic* system in the sense implied by the above definition of the term, except that in our case, such a system is carried by organisms with a central nervous system. We posit that modeling is a selectionist process of evolving mental models, a new component in social systems consisting of interacting agents who perceive, reason, decide and act.

Ademittedly, the idea that modeling leads to learning is not new. Forrester has persistently posited system dynamics as an ongoing process that allows improving our understanding of complex situations: a model is "only a snapshot in time and catch[es] but a single step in a continuously evolving set of ideas about a social system", and that "rather than stressing the single-model concept, it appears that we should stress the process of modeling as a continuing companion to, and tool for, the improvement of judgment and human decision making."

(Forrester, 1985; see also Forrester, 2007). "Modeling as learning" has been and continues to be an actively researched theme in system dynamics (Morecroft and Sterman, 1994; Groesser and Schaffernicht, submitted) and educational research (for a discussion of discovery learning and system dynamics modeling, refer to Schaffernicht, 2010a). Schwaninger and Groesser (2008:449) have stated that a new theory "emerges via variation and selection: options are created, tried out and selected", arguing that in these cases, the (simulation) model "is a theoretical statement". System dynamics takes the closed feedback loop to be the basic construction block of social systems (Forrester, 1996) and assumes that what we believe guides our actions, thus often generating ourselves the problems we later try to solve. Simulation modeling allows learning in and about complex systems from risk-free experimentation (Sterman, 2000: 34).

The idea that a subject who perceives, thinks, decides and acts is part of a larger system consisting of closed loops of causation has been implicit in fields like system dynamics or discovery learning, that have taken little notice of Varela's *oeuvre*. By proposing to conceive modeling as autopoietic system, we show the difference between modeling versus not modeling in a coherent conceptual framework. We are thus able to argue that when modeling becomes part of such a social enactive system, this is a novelty that can be expected to have a selective advantage. A second benefit of our proposition is that rather diverse modeling approaches – amongst them system dynamics – can be a subsystem of such an autopoietic system: despite the differences between them, they can be thought of alternative varieties of the same novelty.

This paper sets out describing an illustrative case where mental models and *enaction* become visible. The following section elaborates an account of selection theory explaining generic processes of acquisition of fit. In this framework, the fourth section proposes models and modeling as evolving, autonomous systems. In the fifth section, we apply this conceptualization to show demonstrate the selective advantage in the illustrative case. We will make special emphases all along the way in special places where the contribution of the ideas of Varela help to characterize models as autonomous systems. We close by underlining that computer modeling enhances our ability to make evolve mental models and thus it develops, and also speeds up, the process of fit between our beliefs and external (and ever changing) referents.

2. A THOUGHT EXPERIMENT CONCERNING POSSIBLE WORLDS

Consider the following thought experiment. Two families –the Araus and the Mendels - live and raise their children in two different countries (A and M), just as many other families. We will look at their ideas (or "mental models" [1]) concerning children, as well as how kids develop ideas about themselves, and how these ideas relate to the way the subjects perceive and act.

No matter which family children are born into, there are no differences between them (lest genetic or other specific problems). We may not know up to which point they have conscious ideas on their mind, but we can safely suppose that they are motivated by two desires:

(i) I want to be like the grown ups;

(ii) I want to make my parents happy.

Thus, their perception regarding their parents' expectations is very sharp (especially including sublingual communication), and their actions are regulated such as to comply with the perceived expectations.

On the parents' side, the ideas concerning children and their raising are different at the outset. So let us first describe the ideas, perceptions and actions of the Arau family (let the parents' names be Alberta and Alfonso, their child's name Alexander). Alberta and Alfonso have certain *premises* (ideas assumed to be true and not open to revision): "children are wild" and "children

[1] In this paper, "mental model" refers to a subject's ideas about what exists, how things are and how they are causally related to each other.

are unable to do the right thing in new situations". They believe this to be true, since it is coherent with their childhood memories, and other parents they know believe the same. Their *perception* is therefore sharpened to "wild behavior" and "new situations". On the side of *actions*, first of all the premises and the according expectation are sub-linguistically expressed. Then there are *action-rules* that state "if wild behavior is perceived, then restrict/punish it" and "if a situation is perceived as new, then give the child instructions".

The dynamics of the Arau family unfolds as follows: Alexander perceives his parents' expectations and will start to behave accordingly – "wild". Alberta and Alfonso *perceive* "wild behavior", *think* "I knew it" and apply the corresponding *action rule*. And when they *perceive* an apparently new situation, they *act* such as to surround Alexander with instructions so that there is hardly any space for personal choices. Alexander *perceives* "restriction" and will obey in his *actions*, because of its desire to make his parents happy. And when given instructions, he will try to follow them, thus reinforcing his parents' premise: if nothing bad happens, this will be attributed to the instructions; and if something bad should happen, it will only confirm that Alexander is unable to behave successfully, thus reinforcing the need to give instructions. Over time, Alexander assimilates that he is wild and more or less unable to do the right thing on its own – just like any normal kid. A "*Menschenbild*" is thus shaped according to which, once Alexander has grown up and become a parent himself, the history will repeat itself.

We now turn to the Mendel family: Mary, Markus and their daughter Melissa. Melissa is just like Alexander at the outset. However, her parents part from quite different *premises*: "children want to be like the grown-ups", "children try to behave well" and "children are bound to learn and are able to behave successfully"; most other families they know believe the same. Accordingly, their *perception* will be sharpened for good behavior (which upon being perceived will reinforce the *premises*) and *action rules* will be "if the child behaves well, be proud" and "if the child behaves successfully, then be proud" or "if some accident happens, reinforce the kid's self-confidence".

Consider now how things develop in this family. Melissa will *perceive* the sub-linguistic expectation and unconsciously try to *act* according to it. Mary and Markus perceive what they were watching out for, think "I knew it" and feel reinforced in their *action rules*. Melissa's behavior will be reinforced towards the premises on behalf of the parents, and she grows into the self-image of being able to do the right thing and to do it right. Once grown up and being a parent, Melissa will do as her parents did and the history will reproduce itself.

The children's "set up" is the same in both cases; nevertheless, the parent's premises and the way they influence their perception and their actions lead their children into behaviors that result in an ever-more stable mental model of children and the way they "must" be treated. As generations pass, there will be quite different predominant ideas about children and child-raising in these two families (and arguably their societies). Two radically different realities have been brought into existence, made up of two radically different sets of mental models, perception rules and action rules on part of parents and children. Beyond the obvious remark that such different families are equally possible, the question is if they are equally desirable for us as external observers and for those who live in them. We might also wonder what a student exchange between A and M or a visit of the Arau parents to M would trigger: would they suspect their own mental models to be suspicious, or would they perceive what they are prepared to perceive and enact an adaptive change in the affected children? Was there a chance for one of the parents to find out that they were actively creating the "reality" they then perceive?

We now have the elements needed to engage in an explanation of how such living systems as ourselves have developed on Earth, and how "modeling as learning" makes it easier to break out of the premises' dictate.

In the thought experiment, there were two agents each time: the parents and the child, each part of the other's milieu. The following figure illustrates this pair of structurally coupled systems:

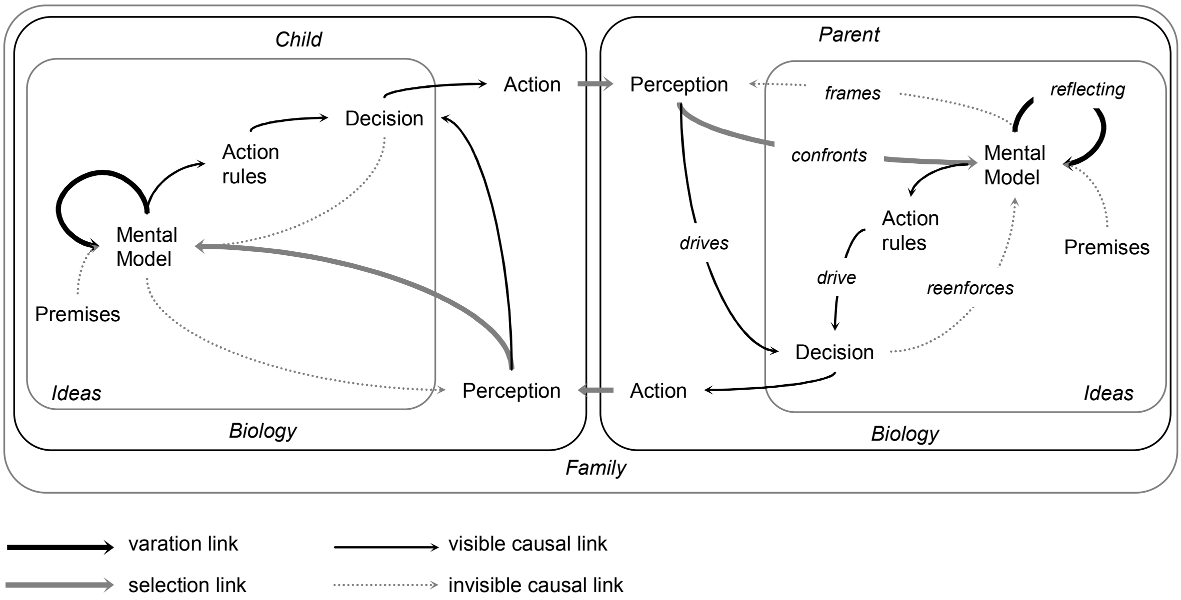

Figure 1: two autonomous and coupled systems

Parents and children alike are depicted as autonomous "family" systems that are structurally coupled to one another since one's action triggers the other one's perception. Each of them is a biological system (solid round rectangles), but also a system of ideas (grey round rectangles): humans cannot avoid having mental models. Reflecting is a process that generates *variation* in form of new alternative (parts of) mental models.

There is an obvious closed circuit of links: one's mental model frames action rules which drive (together with perception) decisions that lead to actions which trigger the other one's perception (and in the other agent, the same process occurs). This framing effect is usually not conscious to the subject, nor is the fact that decisions taken tend to reinforce the mental model that justified them. Only what is perceived by one of these agents can lead to a revision of the mental model (or not, depending on how open the agent is towards "surprise" perceptions): it is the unique opportunity to select an idea or discard it. Since perception needs the other agent's action, the possibility to select depends on an external factor from each agent's viewpoint.

Clearly, parents and child are bound into an interdependency they cannot escape. However, even though each of the connecting arrows suggests that they are aware of these connections, we doubt that they understand the dynamic implication: that which one perceives from the other depends on one's own previous actions.

There are several relevant aspects to note. First, this is an enactive system in the sense of Varela (1995:330): "perception consists of perceptually guided action". The family as a system emerges from the interactions between the components "parent" and "child", which may be labeled a "creative circle" (Varela, 1981): it is in itself an autonomous system, since as part of its usual operation and regenerates the parent-child relationship.

Second, this very aspect is not known to the individual agents inside the system. The dotted lines from decisions and from the premises to the mental model explain a tendency of mental models to be self-concealing; the dotted arrow from mental model to perception indicates that our only chance to sort out flawed ideas – perceiving that they do not work – is endangered by the mental model's framing effect. Ideas lead to new ideas (dotted arrows), but we are not aware of our thinking (Bohm, 1980, 1992); ideas influence the way we perceive and the way we take decisions (as perception and action rules), and decisions in turn influence further ideas. Bohm wrote that "thought creates the world and then says <it was not me>"; even though we do not have proprioception[2] of our cognition, we can learn to pay attention to how we think and how we use words to create meaning.

Third, since there is no trustworthy way to internally sort ideas into "useful" and "discarded", each agent depends on his milieu to indicate his ideas' usefulness. Recognizing errors as such is therefore a slow process.

Fourth, even though these processes are usually not accessible to our awareness, there are approaches to "suspend" these processes and make them available to conscious thought (see Bohm, 1996 on "suspension" or "making visible"). Note that when a subject thinks about his perceptions, decisions or actions, the content of his thoughts are ideas about his perceptions, decisions and actions: we do well to separate the rules for perceiving, deciding and acting into the two classes proposed by Argyris: espoused and in-action.

As we will show in the coming sections, modeling can be used as a suspension device and transfer the locus of selection from external to internal, in the following steps:

1. The individual, being aware of his own existence, is centered on his own perceptions. Therefore networks of ideas (mental models) are a central component of an autopoietical system of perceiving, thinking and acting: even though they cannot exist without the basement of a biological system, they fulfill the conditions that describe autonomous systems.
2. Humans have always developed ideas as part of living; however, simulation modeling ("modeling to learn") is a variation that transfers the selection of ideas from the *external* to the *internal* realm, since ideas can be confronted with their use without passing through other agents.
3. This variation should give a selective advantage to those who incorporate it, for mental errors can be sorted out in little time and without intervening in the unfolding of the enactive interactions. It follows that the "modeling as learning" will diffuse in our population, but it may take generations .

The first point has been brought forward by our thought experiment. We can organize the realm of ideas (Morin, 1991) into dogmas and theories. Dogmas are not open to critique or revision on grounds of experiential evidence; theories are. Consequently, theories evolve over time, while dogmas do not. It cannot be said that dogmas are not viable; the thought experiment showed the process, and the overwhelming stability of organizations based on dogmas would belie any attempt to call it unviable. However, as noted above: viable is not the same as desirable.

Point 2 is the central part of our elaboration. Therefore, we now describe evolution as a generic process, which is a fundamental part of our argumentation, and will then apply it to mental modeling. Following this, we will derive point 3.

3. SELECTION THEORY

In our view, the way we go about perceiving, thinking (producing ideas) and acting – interacting with other systems in the world surrounding us (our milieu) is rooted in the generic schema of evolutionary adaptive systems. Specifically, following the tradition of evolutionary epistemology, we can conceive of such a processes as evolutionary—*selectionist*— processes. This section briefly introduces this selection theory.

Although there are variations on selectionist explanations, the basic tenet of this approach is based on two interacting steps:

(i) *Variations* that provide material upon which selections act on.

[2] the capacity to perceive one's movemens

(ii) *Selection* due to the elimination of unsuccessful forms by the environment or by internal constraints or internal selection processes.

This *variation* + *selection* combination forms an algorithm with substrate neutrality, that is, its logical structure implies guaranteed results independent of the "materials" that happen to be used to carrying it out (Dennett, 1995). This abstractness provides a generic schema that can be instantiated to give an actual theory (Darden & Cain, 1989) by proper theoretical recontextualization in specific domains (Dopfer, 2005; Hayek, 1942). Selection theory is perhaps the most accepted explanation of generic processes of fit, that is, the explanation of satisfaction relationships in which one thing comes to be adapted to another thing (Bickhard & Campbell, 2003; Cziko, 1995; Darden & Cain, 1989). This statement asserts then that it defines an abstract form of explanation.

The most famous instantiation is the neo-darwinian schema that explains the development of life on Earth. But other domains have also been benefited from this approach. For example we currently know that the immune systems of vertebrates function like 'Darwin machines' (as opposed to the classic von Neumann Machine) in which the antigen plays a selective role on a diverse population of antibody molecules (Jerne, 1955, 1967). In neuroscience Gerald Edelman coined the expression 'Neural Darwinism' (Edelman, 1993; Edelman & Tononi, 2000) for picturing the brain as a Darwinian environment where selectional mechanisms drive the formation, adaptation, and interaction of collections of interconnected neuronal groups (Sporns, 1994). These latter ideas have been the main engine to make a paradigmatic change in artificial intelligence (Reeke Jr. & Edelman, 1988) through the construction of "Darwin automata" that have been shown to develop, among other capacities, perceptual categorisation, invariant visual object recognition, forms of learning, and adaptive behaviour (Fleischer & Edelman, 2009; Krichmar & Edelman, 2008). The evolution of human language and the specialisation for grammar have been also explained as a selectionist process (Pinker & Bloom, 1990; Yang, 1999). The list of application domains is long: cognition (Piattelli-Palmarini, 1989), computation (Mitchell & Taylor, 1999), creativity (Campbell, 1987a; Cziko, 1998; Simonton, 1999), memory (McNamara, 1999), cosmology (Smolin, 1992, 1997), educational processes (Perkinson, 1984; Salas & Olaya, 2009), scientific knowledge (Popper, 1972), evolution of technology (Ziman, 2000) and cultural evolution in general (Campbell, 1965; Nelson, 2007). It must be emphasized that all of these cases are neither applications nor analogies of the theory of evolution of living organisms. It is the other way around. They represent different instances of the application of the abstract form of selectionist explanation of fit; the evolution of life is but just one example.

Selectionist explanations are typically grounded on the specific metaphysics of *process* philosophy, e.g. (Heraclitus, ca. 500BC; Rescher, 2008; Whitehead, 1978), as opposed to the common metaphysics of *things*. Thus, it is usually recognized as a suitable way to approach novelty and *emergent* complexity (Olaya, 2007). This type of explanation is not the common causal approach inherited from Hume (1740) that it is still very popular today but, instead, it becomes the explanation of a *process* based on historicity and related *explanantia*, which demands a very different stance in what is understood as a scientific explanation (Mayr, 1969, 1982, 2004). This is an important common point with "biology of cognition" which takes cognition as a process. Inded the schema of organic evolution is an emblematic and paradigmatic process for process philosophy (Rescher, 2008). The idea of "algorithm" (Dennett, 1995) may encompass a large part of this position. As we mentioned, this algorithm is a two-step process; let us take a look at them.

Variation

Selection operates only if variability is available. The process of variation provides the "raw material" for evolutionary processes. This variation should be generated in copious and dependable amounts, so as to form a heterogeneous set characterized by unique individuals (Mayr, 2001).

The three main types of variation processes are: random, blind, and biased. In *random variation* processes, all responses are equally likely due to chance factors that cannot be anticipated or identified; it behaves statistically as a random probability function (uniform distribution) .

A more restricted form of variation is *blind variation,* expression coined by Campbell; “blind” essentially denotes that variations are produced without *a priori* knowledge of which ones, if any, will furnish a selectworthy encounter (Campbell, 1987a); therefore, "blind" does not mean strict equiprobability of alternatives, but the probability distributions of variations are independent of previous experiences, potentially successful trials or goals. Specifically, three conditions are required: (i) idependency of the environmental conditions; (ii) no correlation with the solution—specific correct trials are no more likely to occur at any point in a series of trials than another, nor than specific incorrect trials; (iii) no "correcting" process between variations—that is, a variation subsequent to an incorrect trial is not a "correction" of an earlier one. Thus, this process is not completely random, it does not start from scratch every time anew: it builds on what already exists (Dewitte, 1999). Indeed, blind variation does not exclude the possibility of intrinsic constraints on what is a possible variation, that is, constraints that are *intrinsic* to the ontology of the phenomena under investigation (Bickhard & Campbell, 2003).

Finally, *biased variation* is a corrective goal-seeking process in which the generation of variation is conditioned by externalities, by previous efforts or by previous acquired certainties. Change is directed in generating variations which is a typical characteristic of processes of intended progress. A good example is Beer’s model of organizational cybernetics (Beer, 1966, 1979, 1981) in which variety generators are conditioned from outside so that the mutations of the adapting system are biased according to a particular reward function (Olaya, 2008).

Due to variation, different entities will exist. Because the milieu and the resources are limited, the environment will provide some restrictions for the survival of all of the entities, but some of these will cope better with such restrictions or with the changes in the environment so that this process will necessarily end in the selection of the more adequate variations.

Selection

The direction of evolutionary change is granted in the second step, in the form of *selection*, which works upon variation. The beneficiaries are the entities that are left over after all the less fit entities have been eliminated; this is a *nonrandom* elimination process driven by “selective elimination, selective propagation, selective retention, of certain types of variations, e.g. differential survival of certain mutants in organic evolution, differential reinforcement of certain responses in learning” (Campbell, 1965, p. 27). Furthermore, it is typical to have a “mechanism for the preservation, duplication, or propagation of the positively selected variants” (p. 27), for instance *memory* in learning processes.

If this fitness is determined by relatively stable elements of the environment then it is a case of *natural* selection, the best fit individuals have the greatest probability to “survive”, that is, the best adapted to the momentary environmental constellation (Mayr, 1991). Such a process is opposed to *artificial* selection that implies a deliberate choice made by a knowing decision maker (individual, institutional) that selects based on *a priori* defined criteria; a good example might be the “scientific selection” of the best workmen suited to specific types of work based on criteria of efficiency as emphasized by Taylor (1911) in his influential “principles of scientific management” which are still nowadays the basis for most selection procedures in organizations.

In both cases, selection is *external* to the entity subject to the process. However, there is also *internal* selection for coherence such as the requirements of organizational stability (Bickhard & Campbell, 2003). Internal selectors—or “structural” selectors (Campbell, 1987c)—arise from inner conditions and from the internal organization and activities of the system and do not involve any selection by the environment though they may include representatives of external selectors (Campbell, 1965, 1987b). For instance, in the domain of scientific knowledge evolution, “internal selection” refers to selection for compatibility with the trusted corpus of scientific beliefs and with the social system requirements of the scientific community

(Campbell, 1987c). The internal-external distinction naturally depends on how the boundary of the system is drawn, e.g. the biosphere self-organizes into cells, species, organizations and ecosystems, through selection processes that are internal to the biosphere but that also are external to modular components of this system (Bickhard & Campbell, 2003). The selective systems and constraints are habitually numerous, closely intertwined and, in many cases, very difficult to identify or specify (Campbell, 1987b; Endler & McLellan, 1988).

Replication and self-replication

We want to argue that networks of ideas are self-producing systems. Therefore, we insert a special point about *replication*. Usually the unit of selection should be a replicator, i.e. an entity of which copies are made; this copying process habitually refers to a structure that is passed on largely intact, e.g. in organic evolution it would be the gene (Nanay, 2002). Typically the replication process explains the heritability of variations. Although not necessarily all processes of evolution by selection require that something plays the role of a replicator, the main point is that heredity should be conceived as a correlation between "parents" and "offspfring" independent of how this correlation is achieved; replication is just one possible mechanism (Godfrey-Smith, 2000). Some processes of replication may generate variation, i.e. imperfect fidelity in the copying process; in this latter case, replicators will be slightly different through generations: the changes accumulate and explain therefore adaptation processes (Nanay, 2002); this intrinsic historicity of evolutionary processes explains emergent structures accumulated through time (Bickhard & Campbell, 2003). In particular *self-replication* processes refer to replicators that replicate themselves.

In turn, three types of self-replication can be identified. On the one hand, there are plain processes of *transcription* that replicate descriptions, i.e. self-description (following von Newmann the code is used as "uninterpreted information"), for instance a photocopy of this paper, or the process of meiosis in which the DNA code is copied to produce new identical strands of DNA. On the other hand, there are processes of *translation* that replicate self-assembly instructions to be interpreted so as to be able to build new replicants ("interpreted information"), for instance in successful gametes strands of DNA are decoded to synthesize the proteins to construct a body during development (Sipper, 1998). In addition, it is possible to conceive of *uncoded* self-replication in which self-organized structures emerge from local interactions (as opposed to directed by coded instructions) (Gabora, 2004). We suggest that positive and negative feedback processes may characterize this latter type of replication. The coupled interacting systems of Fig. 1 are a good example, that is, the self-replicating fate of the Araus and the Mendels.

As a summary of this section:, evolution is defined as heritable variation of fitness (Nanay, 2009). Variations are generated and selected through evolutionary cycles and instances of fit are achieved by selection on an abundant generation of possibilities. Given these processes then an evolution in the direction of better fit to the selective systems becomes inevitable (Campbell, 1965). In particular, self-replicating entities may evolve as self-organized structures from local interaction processes; the intercoupled systems of child-parent are examples of the latter. In fact we will later argue that mental models are evolving self-replicators. The next section will explore this type of systems.

4. MENTAL MODELING

Background

There are arguably various possibilities as for naming what is going on in our minds. One may focus on the process of thought (with various sub-classes) or on its product, which can be called thought, idea, belief, knowledge or model. We have chosen to call the process "modeling" and a "model" is part of what is produced by *modeling* (there are also other parts like "action", and additionally a model constrains the modeling process). While doing so, we are aware that according to the "biology of cognition", cognition is a process; however, cognitive processes occur on different time scales and if we are interested in a time horizon adjusted to the frame of awareness of a human subject, some of the processes are comparatively slow and their current state of development (during each of these short periods of time) can safely be regarded as a product. Rather, we want to examine the process of modeling, i.e. the way a mental model constrains action but also the very process of modeling.

Several decades ago, Kenneth Craik (1943) proposed that thinking is the manipulation of internal representations of the world, and the notion of mental models appeared. Some disciplines have picked up the term early on and considered that mental models are a broad class of representations humans make up of something in the world (Forrester, 1961, 1992). Of course, mental models take rather different forms or articulations. As remarked by Forrester (1961), natural language descriptions, diagrams and equations alike are different manifestations of mental models. In this sense, we do not imply that the individuals in our thought experiment mentally construct or manipulate diagrams or equations; however, for instance what they believe can be expressed as discourse or as diagrams. As a typical way to define the notion, they are said to "reflect the beliefs, values, and assumptions that we personally hold, and they underlie our reasons for doing things the way we do" (Maani and Cavana, 2007: 15).

In psychological research, two traditions propose different cognitive mechanisms for studying and explaining human reasoning: rule-based theories and model-based theories (Johnson-Laird, Byrne, & Tabossi, 1989; Knauff, 2007). The former approach is based on the notion of rules of inference; it establishes that the mind is equipped with formal rules of inference that enable deductions. The mental models theory of reasoning asserts that people build models of situations, mental representations that correspond in structure to the situations that they represent (Johnson-Laird, Byrne, & Schaeken, 1994). According to Phillip Johnson-Laird, human beings actually reason and understand the world by means of mental models through the manipulation of abstract mental representations (Johnson-Laird, 2006). In reasoning research, a mental model is a set of logical assertions, interrelated by causal links. When faced with a situation that calls for a decision, subjects invoke one such model for each possibility that may be true and then process them to come to a conclusion.

Recently, evidence seems to favour the model-based view. This theory has continued to be dominant in the psychology of reasoning over the last three decades (Smith, 2008) and it explains several mental operations from which are of interest: creativity, insights (Johnson-Laird, 2006), relational reasoning (Goodwin & Johnson-Laird, 2005), inferences about time relations and temporal reasoning (Schaeken & Johnson-Laird, 2000; Schaeken, Johnson-Laird, & d'Ydewalle, 1996; Vandierendonck & De Vooght, 1996), representation of systems with complex and mixed dynamics (Moray, 1999), detection and resolution of inconsistencies (Johnson-Laird, Girotto, & Legrenzi, 2004; Legrenzi, Girotto, & Johnson-Laird, 2003).

Other disciplines have made reference to mental models, according to their purposes. For instance, in the realm of dynamic systems, "a mental model of a dynamic system is a relatively enduring and accessible, but limited, internal conceptual representation of an external system (historical, existing, or projected) in terms of reinforcing and balancing feedback loops emerging from stock, flow, and auxiliary variables that interact in linear and mostly non-linear, delayed ways, whose structure is analogous to the perceived structure of that system".Groesser and Schaffernicht, submitted, elaborating on an earlier definition by Doyle and Ford, 1998 and 1999)

Following Seel and Blumschein (2009), a mental model can play two major roles: (i) to serve as a concrete, comprehensible, and feasible mental representation of something (e.g., a complex system)—here, the representation of attributes of objects comes second to the representation of structural relationships. (ii) to constitute the fundamental basis for reasoning. In our case, the term *representation* collides with Varela's and Maturana's view that the brain cannot represent something external since no external reference object can be accessed. Therefore, we call them mental *presentations* of problematic situations which are good enough to be useful for the second possibility.

What we have referred to as "ideas" - the components of cognition which play essential roles in human reasoning by serving as building blocks for human thought, are also frequently called concepts (Jonassen, 2006). "Conceptual change" occurs when people change their understanding of the concepts they use and how they are organized within a conceptual framework (Jonassen, 2006), that is, a mental model. This is was system dynamicists call "insight", and we believe it can be detected and measured by adequate comparison (Schaffernicht and Groesser, 2011). This process of conceptual change can be explained as a selectionist process, as it will be introduced next.

Evolution of self-producing mental models

In the domain of cultural evolution, Gabora (2004) introduces the suggestive idea that self-organized, interconnected networks of ideas are uncoded, emergent self-replicators that evolve through time. This will be the starting point to characterize such a process as a type of selectionist development. Indeed, “mental models” develop processes of adjustment, or fit, with their environment. Following Varela, these networks of interactions of components can be labeled as “unities” that belong to a special class of systems - autopoietic organizations- since their components: (i) participate recursively in the same network of productions of components which produce those components, and (ii) realize the network of productions as a unity in the space in which the components exist (Varela, Maturana, & Uribe, 1974). Furthermore, Varela emphasizes that autopoietic systems fundamentally are processes, as opposed to static structural descriptions (Varela, 1997), and thus matching the ground on which evolutionary development is based, i.e. algorithmic processes of continuous variation generation and selective encounters. Relations of processes are highlighted over spatial relations (Varela et al., 1974).

Hence, a selection-type explanation becomes a likely framework for specifying this type of evolutionary growth of fit between mental models and environmental referents, as a subtype of the more general problem of fit (Campbell, 1959, 1987b, 1987c, 1990). So we propose considering modelling as the acquisition of epistemic fit between mental models and their external referents. This is a co-determined process in which fit is achieved by both internal and external processes without assuming perfect fit or perfect representation as in a form of naive or even direct realism (Blackmore, 1979) which was also criticized by Varela (1993). We agree with Campbell (1990) and assume inevitable gaps in the fit of phenomena to nomena. This point will be particularly important for exploring the role of simulation models; learning for us involves vicarious selectors rather than direct encounters with a “reality”. Here knowledge is conjectural, imperfect and co-evolved.

Variation

The generation and production of mental models form the population of systems to be selected. Where does variation come from? First of all, we can think of our mind as essentially variable and changing. McNamara (1999) underlines that “variability is Mind’s central characteristic” (p. 77) and relates this inherent attribute with William James’s idea of the “stream of thought”, i.e. thought is always changing: “now we are seeing, now hearing, now reasoning, now willing, now recollecting, now expecting, now loving, now hating...mental objects associated with these states...can seem to recur. Strictly speaking, however, nothing recurs since all of life is subject to the ravages of time. If a mental content returns a second time, it always does so with a new time tag or date and is, therefore, different than when it first appeared” (p. 82). Specifically, Campbell (1987a) illustrates how *blind* variation is necessary in order to go beyond prescience and what is already known so as to have creative thinking; he illustrates it through wide range of areas and thinkers, including Ashby and his Homeostat, Hebb, Riggs, and Platt on vision, Bain on the psychology of creativity, Souriau on innovations, and Poincaré on mathematical inventions, among many others.

We suggest to visualize this process as continuous self-replication activity. Gabora (2004) proposes an uncoded process of self-replication based on selectionist models of cognitive architecture to explain how the relational structured web of concepts, i.e. the mental model, makes imperfect copies of itself via associative thinking:

> …The capacity for abstract thought arose through the onset of a tendency toward coarser coding; that is, more widely distributed storage and retrieval of memories… Thus more memory locations both (1) participate in the etching of an experience to memory, and (2) provide ingredients for the next instant of experience. Given that the region stored to and searched from at any given instant is wider, and because memory is content addressable, similar items are stored in overlapping regions of conceptual space, and sometimes get retrieved simultaneously…reminding events increase the density of the stored items by triggering the emergence of abstractions (e.g. concepts such as 'cat', 'container' or 'democracy'). Abstractions increase the frequency of reminding events because, via associative pathways, they unite all their instances (e.g. specific experiences of cats). Reminding events themselves begin to evoke reminding events recursively, thus generating streams of associative thought, which increase in both duration and frequency. In the course of these streams of thought yet more abstractions emerge, which themselves become connected in conceptual space through higher-level abstractions… memories and concepts undergo a phase transition to a state where each memory and abstraction is retrievable through a pathway of remindings/ associations. Together they now constitute an autocatalytically closed, relationally structured conceptual architecture, or worldview, that both creates, and is created by, streams of thought (pp. 134-135).

This process of uncoded replication (since there is no a copying process in strict sense but instead influences that are psychologically or culturally attractive) is characterized by low fidelity which guarantees uninterrupted variation and thus, a source for selective systems to act upon. In our thought experiment, and following Gabora (2004), the worldviews of the children are replicants of the worldviews of their parents (and others) and vice versa, e.g. "parental worldviews composed of ideas, attitudes, and so forth that foster the development of a more or less coherent, useful, and satisfying worldview in the child would seem to be at a selective advantage" (p. 137). But this is an imperfect copying process; many associations, modifications, influences, etc. are at disposal. Mental models are shaped by external influences that we relate with our present worldviews in order to build in our mind new further coherent models.

Biased variation processes (for instance the ones that reinforce previous acquired knowledge) favours the preservation of the integrity of the system but fails to go beyond what is already known. On the other hand, *blind* variation guarantees true innovation which is necessary for survival in changing environments (as biology shows). For instance, maybe one day, Melissa earlier-child and now teen-ager will challenge her parents' expectations ("children want to be good") and perhaps she will start to behave very differently; probably external selective pressures put on her by her friends will explain continuous changes (variations) in her mental models; perhaps first only a rare trial will be executed (e.g. one night she arrives from a party later than the allowed time by her parents); as time goes by, and surely depending on the results, some of these attractive variations will be maintained and selected.

On the whole, the mental models of Melissa will be transformed because of her exposure to numerous external influences; these mental models will have resemblance with their ancestors but the constant low fidelity self-replication of her worldviews, with variations introduced as experience and thought goes by, will guarantee that her worldview changes through her life. Thus, in the same way that Darwin showed for living organisms, the key to have successful adaptive (belief) systems relies on free (blind) variation.

Selection

Naturally there is elimination of ineffective mental models and temporal retention of successful ones.

Regarding internal selectors, McNamara (1999) compiles experimental evidence of mechanisms that inhibit or suppress irrelevant or distracting elements that facilitate selection of target systems which, once selected, are entered into long-term memory. He also shows that these selective mechanisms match neurophsychological models such as the above mentioned "Neural Darwinism" of Edelman. Numerous higher level internal selectors are possible, for example Campbell (1987a; 1991) underlines the reasoning of Poincaré on mathematical creative thought in which "mathematical beauty" is the vicarious selector which screens out 99% of the dross produced by the unconscious wild permutation of ideas.

Campbell (1990) also proposes temporal decision-rule structures that form criteria, action rules, and recipes for behaviour that assume stability and external regularities; he also portrays distortion-correctors that actively construct and make sense of very imperfect perceptions and thus improving belief although in a very indirect way, e.g. motion pictures perception. Other forms of internal selection can be triggered by external pressure, e.g. selection via criticism as in hypothetical realism (Popper, 1972); processes of reflection (which attempt to anticipate future external selectors) or social argumentation and counter-argumentation may make us eliminate "bad" ideas in our mind (Salas & Olaya, 2010). "Validation" with experts is a good example. Numerous selective systems can be imagined.

How the system maintains its identity in spite of external selection? We want to focus in particular on the ability of the system to maintain autonomy based on the balance between internal and external selection. It should be noted that perhaps the first conciliatory point between self-organizing complexity and external selection was made by von Bertalanffy, this point is taken by Campbell (1991) to emphasize that both processes are needed, that is, autopoiesis and selection. In particular autopoietic organization provides more complex units to be selected from and, more important, the autonomy for an internal organization to keep going through adapting processes. Campbell identifies Poincaré as perhaps the first thinker that imagined the cognitive process in autopoietic terms: "Ideas rose in crowds; I felt them colllide until pairs interlocked, so to speak, making a stable combination" (p. 168).

In particular, since the described cognitive operations work by association, mental models provide a domain of interactions that defines what is significant for the cognitive system - a point of reference. This latter characteristic is underlined by Varela (1997) and it is perhaps what best explains internal selection: the cognitive system attempts to fill the lack of signification against the world by attempting to re-establish a coupling with the environment (*action*); what is meaningful is retained in the history of the organism. Such continuous internal selection assures coherence and also demarcates and separates the system from the environment. This process of continuously bringing forth significance is a source of autonomy for the system (Varela, 1997). This point also explains what co-definition or co-evolution means here: on the one hand there is external selection, independent selectors that act on organized systems (like asteroids on dinosaurs) and, on the other hand, there is prominence of actions directed to what is "relevant" for the knower, the development of meaningful environments (in the same way that a species actively constructs its own ecological niche).

We have shown that the product of the cognitive operations that drive the evolution of mental models is the system itself, i.e. mental models that produce mental models, and thus, this continuity guarantees the autonomy of the cognitive system since it compensates for perturbations through the realizations of its operations. Furthermore, the cognitive results are by definition relational, systemic, associations of ideas. Mental models are autopoietic. And the realization of their autopoiesis can be described as a selectionist process in which low fidelity self-replication guarantees adaptation to changing environments. We will now focus in the special activity of modelling and will expand it to mental, formal and simulation models to sketch the way we understand "learning through modelling" .

5. SIMULATED MODELS AS INTERNAL SELECTORS

Modeling is the activity of giving form to a (mental) model. Nobody can know where an individual's mental model is unless the person articulates it (in his consciousness or talking aloud) [3]; this is a first act of modeling and in principle the very act to articulate can have an influence on the model. The trigger for such initial steps can be external (that which everybody knows or more specific external sources) or internal (Kahneman's "system 2", 2002). If a dogmatic stance is taken, there will be no more significant changes to the mental model. As long as the individual stays open to new perceptions, new evidences will be the opportunity to make changes to the model. In other words, *selecting* failed ideas out of the mental model initiates the generation of *variation*. This leads to the proposition that a model is not a static object, but evolves over a sequence of versions (Schaffernicht, 2006).

'Modeling for learning' is then a deliberate process of producing experiences and evidence that allows revising and correcting the model, de-selecting ideas recognized as not useful. Schaffernicht (2010a) describes the iterative steps of developing a mental model by trying to develop a simulation model that represents it; in this case, the simulation model is used *in lieu* of the external world in order to try the mental model out.

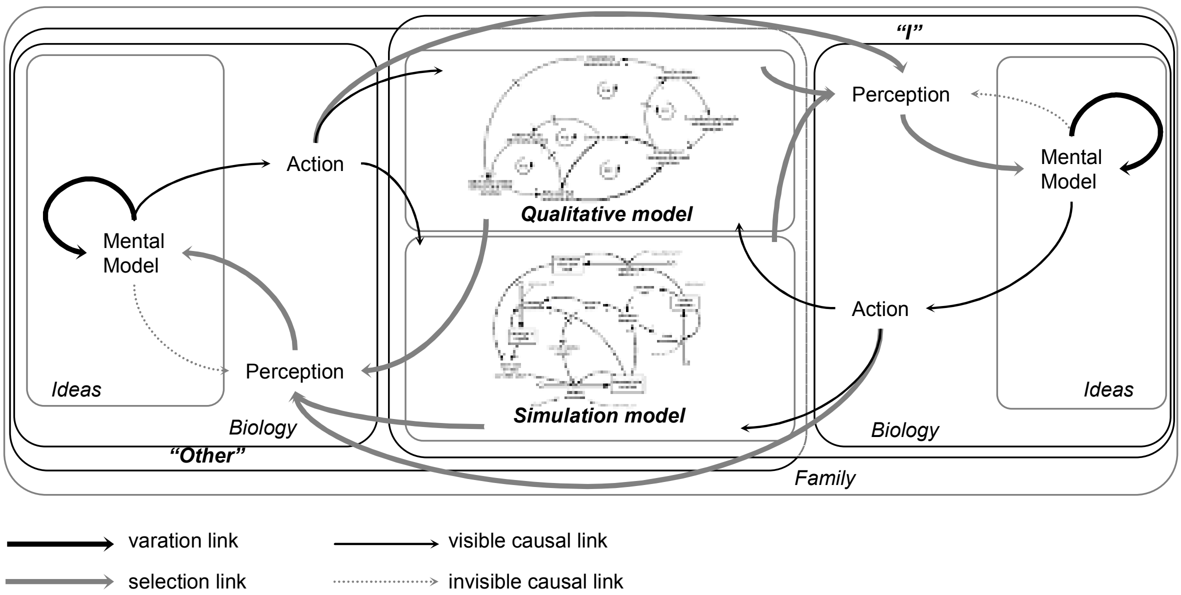

Figure 2: the simulation model as internal selector

This process is presented in Figure 2. In this figure, it becomes apparent that there are two types of models and that each brings a new channel for action and for perception for each of the two agents. A "qualitative model" is one that only states variables, causal links and – in the case of system dynamics - feedback loops. Such "causal loop diagrams" have distinctive advantages and shortcomings (Lane, 2008); they are useful for problem structuring, since the modeler explicates his causal beliefs, and by looking at the diagram he can perceive his "suspended" ideas. This allows to detect conceptualization errors (*selection*) and to generate better conceptualizations (*variation*) of the causal structure of the situation. The behavioral characteristics of the explicated causal structure are generated by the "simulation model", which is a quantified and detailed version of the same causal structure. It allows the modeler to act - formulate and simulate change scenarios (generate *variation*) and perceive the behavior that would be the logical consequence; "surprise behaviors" (Mass, 1991) trigger new reflection, the identification of errors (*selection*) and ideas (*variation*). Developing both types of models thus

[3] Not even the individual himself.

constitutes two additional loops of "perceptually guided action" (further described in Schaffernicht, 2010b).

It is important to note that both models are "internal" to the agent's personal sphere, even if they do not physically belong to his biological organism. According to the 'tacit knowledge' perspective, becoming a modeler means that the external device (diagram) is cognitively incorporated into the knowing person (Polanyi, 1983). The modeler perceives the (external) situation through them, without needing the external loops (running across the other agent) to reflect (generate *variation*) and debug (*select*). Therefore it is justified to call this "internal selector".

Still, the models (diagrams, equations, simulated behaviors) as objects are external to the agent. Consequently, the models belong to both agents' personal sphere. Where both had only action/perception before, now they can jointly develop such models, refer to them in their communication and other actions.

Comparison of figures 1 and 2 clearly shows that with modeling, the social system was already autopoietic. However, we propose that the system described by Figure 2 has a selective advantage. We see two reasons for this.

First, as has long been stated, simulation allows to compress time and space, and that simulation modeling is a useful activity for "learning in and about complex systems" (Sterman, 2000). To these statements, we now add a new component: the generation of new candidate mental models is the "variation" phase in a process of evolution. Therefore, a person can come up with a variety of mental models of worlds that could be. We showed above the importance of this process to be *blind* and not only *biased*. The "selection" phase is then taken care of by the milieu: if actions derived from the mental model prove to lead to unviable effects, the mental model is shown to be flawed (and so the person has to think anew, either correcting the mental model or adding the necessary elements to blame external factors for the failure). In terms of selection theory, this is an "external selector". In this case, the possibility to develop a simulation model transfers the locus of selection from external to internal: the subject can use the simulation model as a kind of a mental extension and sort out errors in the model by "internal selection". Campbell would say that *the simulation model becomes an imperfect vicarious proxy for external selection requirements*.

Second, the systemic intersection between the agents generated by the models and their development generates a communication space. In cases like the "Post-mortem assessments for disputes and learning" (Lyneis and Ford, 2007:170), the very fact to explicate each agent's beliefs enables an iterative process that eventually ends with an agreement, thus avoiding more expensive and damaging outcomes.

The first aspect enables to learn in little time, with little risk and relatively little cost – without having to learn from failed implementation. The second aspect improves the quality of communication between agents, avoiding the costs of disagreements. For both reasons, we propose that social systems with agents who model have the selective advantage of learning faster and at lower costs, also avoiding the costs of insufficient mutual understanding (communication).

Of course, we have to remember that the final test of a model is if action derived from it works out in 'real life' (Forrester, 1961). We interpret the enactive relationship between the agent and its milieu in the very same way: the internal selection carried out by the quest for a 'valid' model does not replace the external selection, but it partly prevents the need for it. Thus it has to be assumed that modeling does not always avoid bumping against a problem.

6. MODELING: A SELECTIONIST PROCESS

If the possibility to have an internal selector is an advantage, then this quality should be apparent in our thought experiment. So consider what a visit of the Arau parents to the Mendel family would lead to.

The default scenario assumes that both families have a dogmatic stance towards their mental models. Therefore the Araus would disapprove of the Mendels' permissive attitude, and the Mendels would not understand this criticism. Each would cling on to their respective mental models.

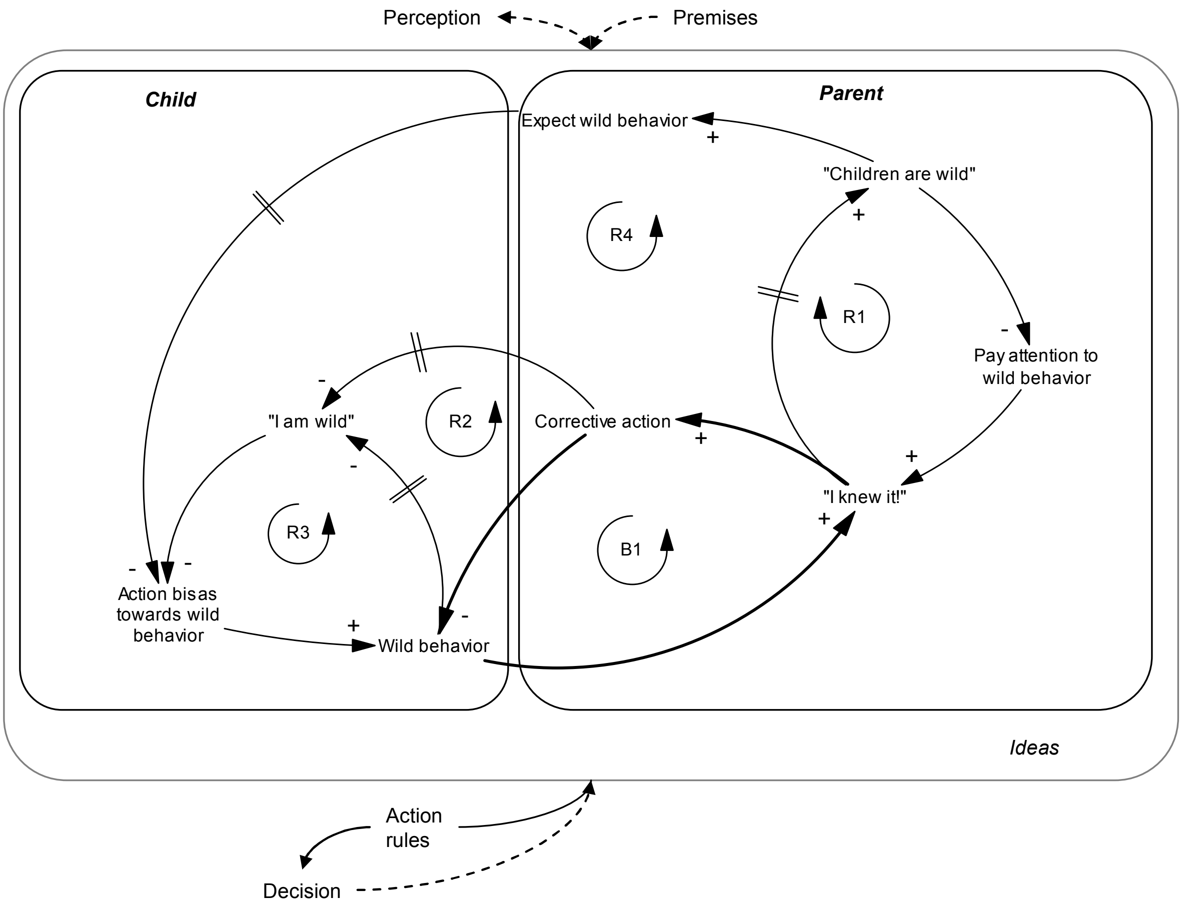

Figure 3: the Araus' mental model inside its context of perceiving and acting

Figure 4 expresses the Araus' mental model; it contains the mental model of the two agents and a complex of several closed loops, represented as a "causal loop diagram" in the framework of the two structurally coupled agents. These diagrams are useful for representing and organizing a system's causal structure with special emphasis on closed loops (Lane, 2008; Schaffernicht, 2010b). The most salient (easy to perceive) loop is B1: "children are wild", therefore it is natural to expect wild behavior and necessary to pay attention to it. When perceived ("I knew it!"), it only confirms what they knew and corrective action reduces it (a balancing or "counteracting" loop). The loop R1 might be in sight, but it operates slowly: due to the delay between the "I knew it" thought and the idea that "children are wild", this influence is barely perceptible, so the loop's self-reinforcing operation may go unnoticed.

The dotted arrows represent influences that might be mentioned by the Mendels in order to explain their astonishment of how bad the Araus think children are. Another delayed link states that repeated corrective action will foster the child to believe "I am wild"; the resultimg bias towards wild behavior will not fail to trigger ever more corrective action (loop R2). Additionally, the very fact of behaving wildly reinforces the idea that "I am wild", which leads to loop R3. Last not least, the idea that "children are wild" leads to tacitly expressing wild behavior, which also reinforces the action bias towards such behavior.

The Araus may recognize them or not, and then the Mendels would explain their point of view (Figure 5).

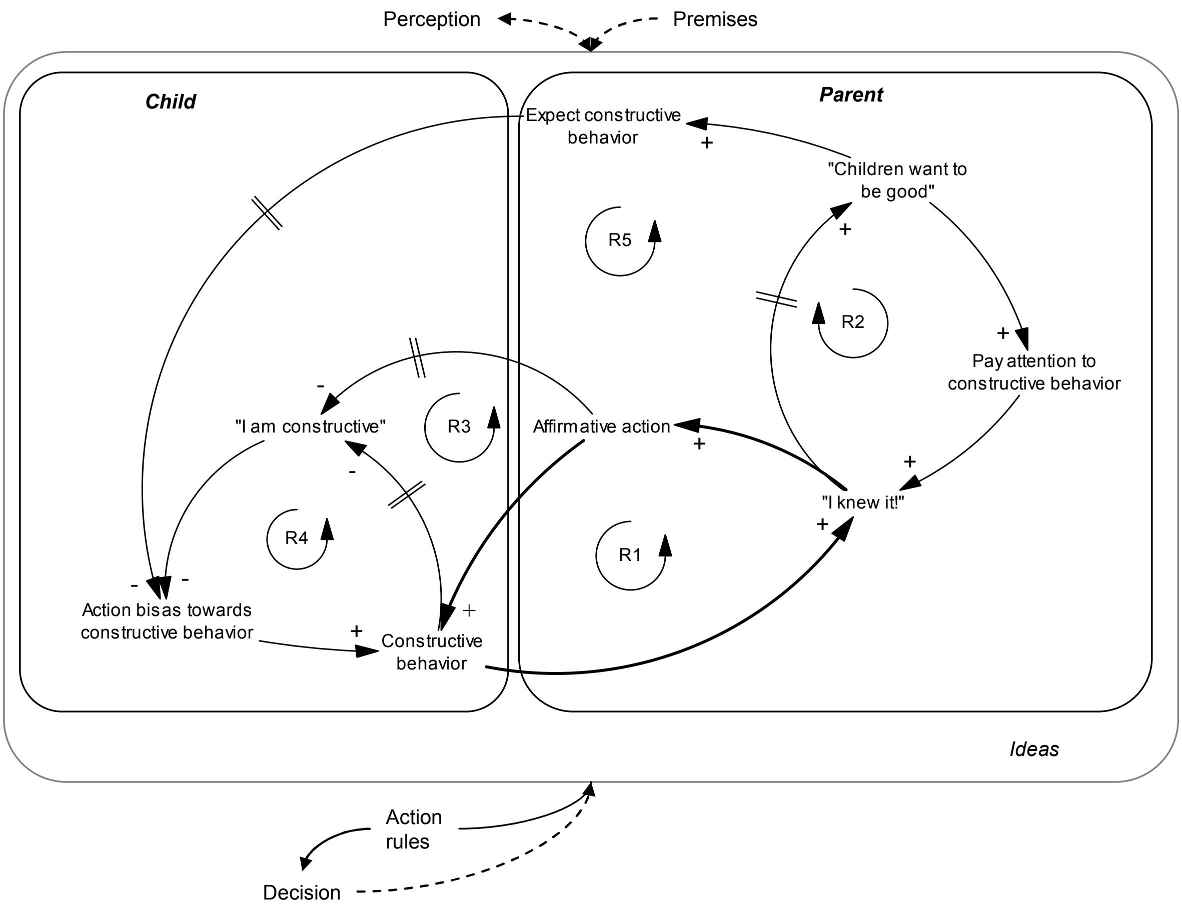

Figure 4: Medel's mental model inside its context of perceiving and acting

“Children want to be good”, so it is logical to pay attention to behaviors that confirm this view. The unavoidable "I knew it" experience leads to a reinforcing loop (R1) of affirmative action, more constructive behavior and "I knew it". Slowly, it also reinforces the idea that "children are good". As before, affirmative action and the "I am constructive" reinforce the action bias towards constructive behavior and the very behaviors (loops R3 and R4). Loop R5 is analog to the Araus' mental model, even though it plays out in a virtuous way.

The dotted lines may not have been perceived by them before, but as they compare the two families, they come to mind. Some debate would take place between these apparently colliding views, and then a new mental model coherent with both families’ experience would become visible (Figure 6).

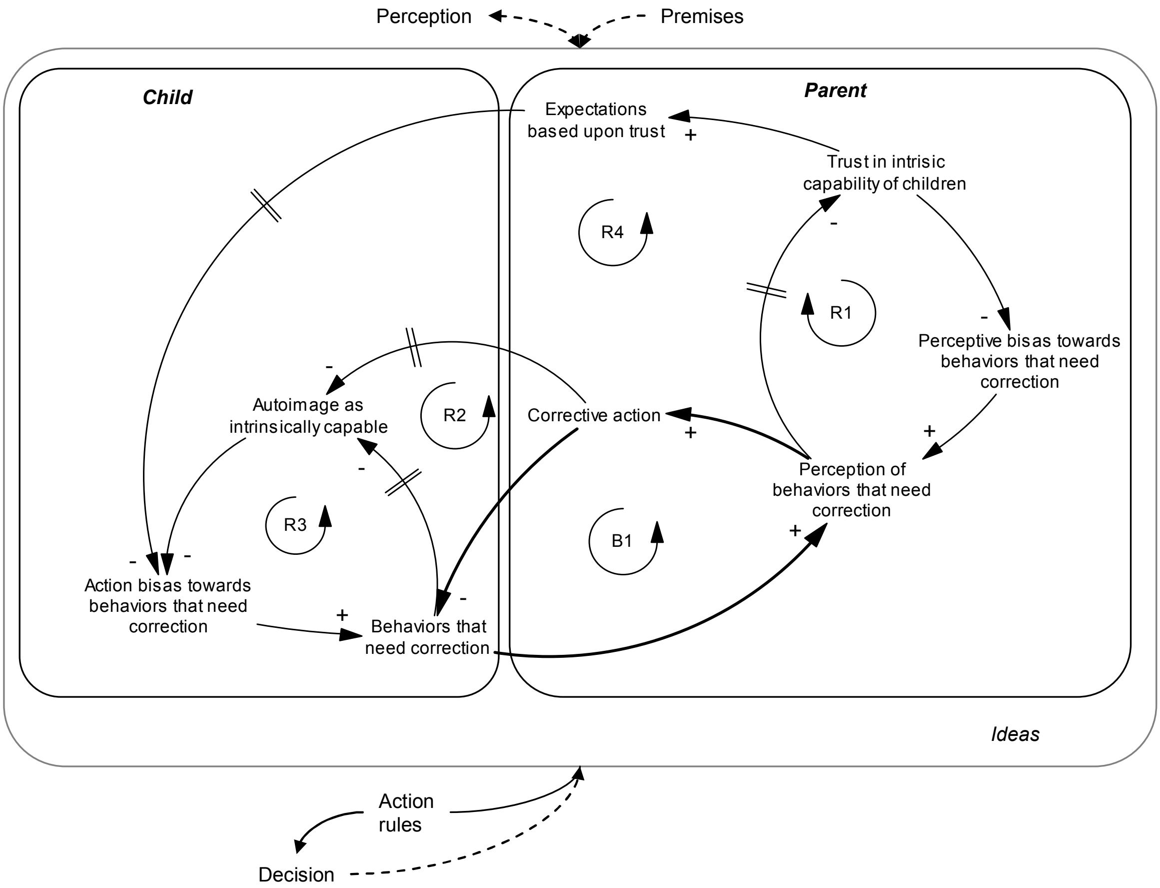

Figure 5: a revised mental model

After their dialectic exchange, the families may now have elaborated a model that reconciliates both families' ideas with a common causal structure. As before, the most salient is B1 (thick arrows), the one that encloses corrective action: the perception of behaviors that need correction (like “wild” behavior) leads to corrective action which diminishes behaviors that need correction. This loop binds parents and children together.

However, the intended correction is counteracted by reinforcing loops that are less perceptible. First, corrective action slowly lowers the auto-image as intrinsically capable of behaving well (R1). This process augments the action bias towards “bad” behaviors and eventually there will be more such behaviors (calling for more corrective action). At the same time, behaving bad slowly lowers the auto-image as intrinsically capable (R2). Also, the perception of “bad” behaviors slowly lowers the trust in the intrinsic capability of children, which leads to augment the perceptive bias (R3) and to lower expectations (R4) that augment the action bias towards “bad” behaviors. Even though the previous figure concentrates on the “children are wild” aspect, it is easy to construct an analogous model for the “children are unable to behave successfully in new situations”.

This new mental model has some important features. First, it overcomes the impression that the Araus and the Medels are acting in different settings. Even though they perceived their worlds to be very distinct, the underlying structure is one, and it allows different stories or worlds to unfold. As compared to Varela’s creative circles, there is a fundamental difference: the system “family” that emerges out of the interdependent actions of parents and children and each of its components are a closed loop of one kind. The closed logical (or causal) loops in our mental model diagram represent processes going on at the level of the components, not the family. However, since all of us belong to the domain of the components of such systems as a family, we can only directly interact with components at the same level. Therefore the feedback loops used in modeling languages like system dynamics do two things:

- they squeeze the "system" into one level of description;
- they allow the agents to figure out what they can usefully undertake at their level of existence.

Having worked out this model, it is now possible for each of the parents in these families to understand that the world is as they perceive it to be because their own actions have generated it this way. Of course, each can decide not to change their behavior – the mental model shows that this is indeed possible, but that there is also an alternative.

Up to this point, no simulation modeling has been involved in the development. Of course, developing such a model is helpful for disciplining the minds' creativity. As a matter of fact, such a model has been developed and shows that in this structure, there is the potential for two radically different development paths. However, describing the simulation model would take space but would not add new information; therefore the model is only provided in the supplementary material.

As recounted, the families had to meet and articulate and confront their diverging views in order to construct the new model; to each of the families, the other family acted like a selector. It is interesting to reflect upon if it is an externals elector or an internal one. If one of the families decides to take the risk and act differently, the events they will perceive will indicate if it works and in this case, doubtlessly the external milieu is an external selector. However, if another agent lends himself as devil's advocate, trying to find the weak point in one's mental model, this is already a kind of selector involving an agent different from "I". In epochs when potent computers were not available yet, this was the usual way to simulate - for instance during the Apollo 13 crisis (Redorbit, 2010). So there is not a fundamental difference between using a person or a computer program as selector of impossible ideas: this is much closer to the agent than the external milieu.

Therefore it is not always necessary to become a modeling specialist capable of handling sophisticated computer software or mathematical methods. However, any of the parents could now infer that models are helpful and that modeling is a worthwhile activity and invest the time and effort to learn simulation modeling or find someone to help them. The fundamental point is that due to the internalization of the selector, errors in mental models can be found earlier and at a lower cost. The "surprise" effect and the unfreezing of mental structures in response to it will then occur earlier and more frequently and it can be alleged that thanks to this effect, the modeling individual will elaborate better mental models that will lead to more successful actions. If so, it should be observable that other individuals imitate the strategy of modeling. Alas, this seems not to be the case.

As recounted by Fisher (2005), even though Forrester could persuade managers of large companies to adopt his recommendations, they would not accept that this was due to his modeling. His conclusion was that "I don't really expect to convert them. The only option is to outlive them". In this sense, there is no convincing reason to expect that modeling will diffuse over the human population rapidly, just because it improves mental models and allows more successful actions; in other words, this is likely not to be a Lamarckian process. Rather, to the extent that simulation modeling leads to more success (or avoids more problems) – and that this effect is stronger than other ways to increase success – those who practice it may have a selective advantage over those who do not and tend to be in better positions in society which would give them the human equivalent to reproductive advantage; then the mental model of simulation modeling would spread in the rhythm of generations.

There is anecdotic evidence that the diffusion of new ideas, be they autopoiesis, Montessori education or system dynamics modeling, is a very slow process, and we can conjecture that Forrester's statement about "outliving them" reveals this idea.

7. CONCLUSION

Based upon the above argumentation, we propose that mental modeling is an autopoietic system, organized (defined as a unity) as a network of ideas and perceptions/actions which:

(i) through their interactions and transformations continuously regenerate and realize the network of ideas that produced them; and

(ii) constitute it (the model) as a concrete unity in the domain of ideas by specifying the conceptual domain of its realization as such a network.

(iii) Generate a border defining the network such as to operate as a logically closed unity separated from its environment and

(iv) Configure a flow of ideas which, upon entry into the network become components of the network, and upon ceasing their participation in the network's operation, become part of the environment again.

Clearly, ideas are not to be thought of without the biological substrate described as nervous system (as part of an organism which is structurally coupled with entities in its environment). Since ideas cannot exist without this biological system, there seems to be a crucial difference. A biological machine like a cell is called autopoietic because it literally produces itself as the organized network of molecular relations as long as it lives. If it cannot be said that that ideas produce the brain that thinks them, too, then the denomination "autopoietic" imay be misleading. However, this may be seen as a problem of the relationships between the domains of description: the domain of ideas depends of the biological domain, so in a certain way the domain of ideas is not entirely self-contained. Then again, it a long range perspective, the development of the systems in the biological domain is not independent from the domain of ideas, either. So we postulate the existence of a "creative circle" (in Varela's terms).

Our argument is but a small step beyond the credo of disciplines like system dynamics that have been developed as a tool "for the improvement of judgment and human decision making" (Forrester, 1985). Our suggestion that the crucial advantage of modeling consists in the transfer of selection from external to internal, is only new insofar as it explicitly puts selection theory to work. Thus, we gain the possibility to frame further research, as will be mentioned below.

A second small step is made by showing that mental models can indeed be thought of as central piece of an autonomous, even autopoietic, system. In this case, much of the accusations that these disciplines take a "positivist" or "realist" stance and thus belong to the "hard systems" movement (which is usually not meant as a compliment) appear to be irrelevant. As seen from the enactive perspective, it is not useful to ask if whatever one interacts with in the milieu is objectively real or not. It does not matter, for as long as a mental model allows taking decisions that work out in a viable manner, the conceptual entities referred to in the model may be taken as real, since they are real to "me". It is more useful to wonder how to develop these models in ways that bring about a desirable world in the future.

Our analysis is logically coherent and its fundaments are anchored in the literature. However, it is clear that more research is required before becoming more affirmative. Such investigations can strive to establish the evolution of models over time; have a historical character or accompany endeavors in real –time, either way they will have to span large periods of time. For instance, the model of the national economy developed by and under the direction of Jay Forrester, or the urban dynamics model and the world dynamics model by the same author: how did these ideas evolve in the modelers' minds, and how did they fare in the addressed people's minds?

At a much shorter time scale, it is worthwhile to ask how a modeler generates ideas when faced with "surprise behavior" (Mass, 1991). Varela has advocated for a compromise between the first person and the third person perspective: can we accompany a modeler and record what is going on in his mind in such a way to determine if *blind variation* is at play?

It is a common place to think that self-organization and darwnian selection are unrelated or even mutually exclusive. Although various authors have underlined the necessary complementarity of both types of processes (e.g. Batten, Salthe and Boschetti, 2008), the stimulating ideas of Francisco Varela represent an excellent opportunity for underscoring such assessment. This paper opens a possibility. In our view, the autopoietic organization of a particular system is a process taking place at a very short time scale as compared to the evolutionary process, to which it is one particular object. Since a blind process like evolution can hardly be thought of as an agent that designs an autonomous system, autopoiesis may be the only alternative. As we have argued, in the case of mental models, the relationship is fruitful: if a mental model is regarded as autopoietic system, then the deliberate construction of "simulatable" models makes sense as a strategy to accelerate evolution by internalizing selection and by generating a structure that can be transferred and thus avoids the need to entirely reconstruct it.

We close borrowing from Luis Pasteur: "fortune favors the prepared mind". We hope to have stimulated readers towards asking these questions and undertaking such studies, to the advancement of our understanding on the development of ideas as autonomous systems. And being aware that the variation of ideas that are our proposition is only a logical possibility, we hope it passes the readers' selection and fosters further dialogue.

REFERENCES

Batten, D., Salthe, S., & Boschetti F. (2008). Visions of Evolution: Self-organization Proposes What Natural Selection Disposes. Biological Theory, 3, 17-29.

Beer, S. (1966). Decision and Control. Chichester: John Wiley & Sons.

Beer, S. (1979). The Heart of Enterprise. Chichester: John Wiley & Sons.

Beer, S. (1981). Brain of the Firm. Chichester: John Wiley & Sons.

Bohm, D. (1980). Wholeness and the Implicate Order, London: Routledge

Bohm, D. (1992). Thought as a System (transcript of seminar held in Ojai, California, from 30 November to 2 December 1990), London: Routledge

Bohm, D. (1996). On Dialogue. editor Lee Nichol. London: Routledge,

Bickhard, M. H., & Campbell, D. T. (2003). Variations In Variation And Selection:The Ubiquity Of The Variation-And-Selective-Retention Ratchet In Emergent Organizational Complexity. Foundations of Science, 8, 215-282.

Blackmore, J. (1979). On the Inverted Use of the Terms 'Realism' and 'Idealism' among Scientists and Historians of Science. The British Journal for the Philosophy of Science, 30(2), 125-134.

Campbell, D. T. (1959). Methodological suggestions from a comparative psychology of knowledge processes. Inquiry, 2, 152-182.

Campbell, D. T. (1965). Variation and selective retention in socio-cultural evolution. In H. R. Barringer, G. I. Blanksten & R. W. Mack (Eds.), Social change in developing areas: A reinterpretation of evolutionary theory (pp. 19-49). Cambridge, Mass: Schenkman.

Campbell, D. T. (1987a). Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes. In G. Radnitzky & I. Bartley, W.W. (Eds.), Evolutionary Epistemology, Rationality, and the Sociology of Knowledge (pp. 91-114). La Salle, IL, USA: Open Court.

Campbell, D. T. (1987b). Evolutionary Epistemology. In G. Radnitzky & I. Bartley, W.W. (Eds.), Evolutionary Epistemology, Rationality, and the Sociology of Knowledge (pp. 47-89). La Salle, IL, USA: Open Court.

Campbell, D. T. (1987c). Neurological embodiments of belief and the gaps in the fit of phenomena to noumena. In A. Shimony (Ed.), Naturalistic Epistemology (pp. 165-192). Dordrecht: D. Reidel Publishing Company.

Campbell, D. T. (1990). Levels of Organization, Downward Causation, and the Selection-Theory Approach to Evolutionary Epistemology. In G. Greenberg & E. Tobach (Eds.), Theories of the Evolution of Knowing (pp. 1-17). Hillsdale, NJ: Lawrence Erlbaum Associates.

Campbell, D. T. (1991). Autopoietic Evolutionary Epistemology and Internal Selection. Journal of Social and Biological Structures, 14(2), 166-173.

Craik, K. (1943). The Nature of Explanation. Cambridge: Cambridge University Press.

Cziko, G. (1995). Without Miracles. Universal Selection Theory and the Second Darwinian Revolution. Cambridge, MA, USA: MIT Press (A Bradford Book).

Cziko, G. (1998). From Blind to Creative: In Defense of Donald Campbell's Selectionist Theory of Human Creativity. Journal of Creative Behavior, 32(3), 192-209.

Darden, L., & Cain, J. A. (1989). Selection Type Theories. Philosophy of Science, 56, 106-129.

Dennett, D. C. (1995). Darwin's Dangerous Idea. The Sciences(May/June), 34-40.

Dewitte, S. (1999). Behavioral Variation: A Neglected Aspect in Selectionist Thinking. Behavior and Philosophy, 27, 127-145.

Dopfer, K. (2005). The Evolutionary Foundations of Economics Cambridge, UK: Cambridge University Press.

Doyle JK, Ford DN. 1998. Mental Models Concepts for System Dynamics Research. System Dynamics Review 14 : 3-29

Doyle JK, Ford DN. 1999. Mental Models Concepts Revisited : Some Clarifications and a Reply to Lane. System Dynamics Review 15 : 411-415

Edelman, G. M. (1993). Neural Darwinism: Selection and Reentrant Signaling in Higher Brain Function. Neuron, 10(2), 115-125.

Edelman, G. M., & Tononi, G. (2000). A Universe of Consciousness. In. New York, NY, USA: Basic Books.

Endler, J. A., & McLellan, T. (1988). The Processes of Evolution: Toward a Newer Synthesis. Annual Review of Ecology and Systematics, 19, 395-421.

Fisher, L. (2005) The prophet of unintended consequences. Features: the creative mind. Strategy+Business 40, Autumn 2005. (www.strategy-business.com/media/file/sb40_05308.pdf; 1/31/2011)

Fleischaker, G.R. (1988). Autopoiesis: the status of its system logic, BioSystems, 22, 37-49

Fleischer, J. G., & Edelman, G. M. (2009). Brain-Based Devices. An Embodied Approach to Linking Nervous System Structure and Function to Behavior. IEEE Robotics & Automation, 16(3), 33-41.

Forrester, J., 1985. The "model versus a modeling process", System Dynamics Review 1: 133-134

Forrester, J.W., 1961. Industrial Dynamics. Productivity Press, Cambridge, MA.

Forrester, J., 1969. *Principles of systems*, MIT Press

Forrester, J.W., 1992. Policies, decisions and information-sources for modeling. European Journal of Operational Research 59 (1), 42–63.

Forrester, J. 2007. System dynamics – the next 50 years. *System Dynamics Review* **23**(2/3): 359–370

Gabora, L. (2004). Ideas are not replicators but minds are. Biology and Philosophy, 19, 127-143.

Godfrey-Smith, P. (2000). The Replicator in Retrospect. Biology and Philosophy, 15, 403-423.

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning About Relations. Psychological Review, 112(2), 468-493.

Groesser, S. and Schaffernicht, M. (submitted). Mental models of dynamical systems: taking stock and looking ahead

Hayek, F. A. v. (1942). Scientism and the study of society (Part I). Economica, New Series, 9(35), 267-291.

Heraclitus. (ca. 500BC). Fragments (2001 Ed.). New York, NY: Penguin Books.

Hume, D. (1740). A Treatise of Human Nature (2000 edition). Oxford: Oxford University Press.

Jerne, N. K. (1955). The Natural-Selection Theory of Antibody Formation. Proceedings of the National Academy of Sciences of the United States of America, 41(11), 849-857.

Jerne, N. K. (1967). Antibodies and learning: Selection versus instruction. In G. C. Quarton, T. Melnechuk & F. O. Schmitt (Eds.), The Neurosciences: A Study Program (pp. 200-205). New York, NY: Rockefeller University Press.

Johnson-Laird, P. N. (2006). How We Reason. Oxford: Oxford University Press.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1994). Why Models Rather Than Rules Give a Better Account of Prepositional Reasoning: A Reply to Bonatti and to O'Brien, Braine, and Yang. Psychological Review, 101(4), 734-739.

Johnson-Laird, P. N., Byrne, R. M. J., & Tabossi, P. (1989). Reasoning by Model: The Case of Multiple Quantification. Psychological Review, 96(4), 658-673.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning From Inconsistency to Consistency. Psychological Review, 111(3), 640-661.

Jonassen, D. H. (2006). On the Role of Concepts in Learning and Instructional Design. Educational Technology Research and Development, 54(2), 177-196.

Kahnemann, D. 2002. Maps of bounded rationality: a perspective on intuitive judgment and choice, Nober Prize Lecture, December 8, 2002

Knauff, M. (2007). How Our Brains Reason Logically. Topoi, 26, 19-36.

Krichmar, J. L., & Edelman, G. M. (2008). Design Principles and Constraints Underlying the Construction of Brain-Based Devices. In Neural Information Processing (Vol. 4985/2008, pp. 157-166). Berlin / Heidelberg: Springer

Lane, D., 2008. The Emergence and Use of Diagramming in System Dynamics: A Critical Account, Systems Research and Behavioral Science 25: 3-23

Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (2003). Models of Consistency. Psychological Science, 14(2), 131-137.

Lyneis, JM. and Ford, DN., 2007. System dynamics applied to project management: a survey, assessment, and directions for future research, *System Dynamics Review* **23**(2/3), (Summer/Fall 2007): 157–189

Maani, K.E., Cavana, R.Y., 2007. Systems Thinking, softcover ed.. System Dynamics: Managing Change and Complexity Pearson Education, New Zealand.

Maturana, H. 1975. The organization of the living: a theory of the living organization, Journal of Man Machine Studies 7:313-332

Maturana H. 1980. Man and society. In Autopoietic systems in the Social Sciences, BenselerF, HejlP, KockW (eds). Elsevier–NorthHolland: New York; 11–31.

Maturana H. 1988a. Reality: the search for objectivity or the quest for a compelling argument.Irish Journal of Psychology 9: 25–82.

Maturana, H. and Varela, F. 1980. Autopoiesis and cognition: the realization of the living. BSPS 42, D. Reidel, Boston

Maturana, H.R., & Varela, F.J. (1987), The Tree of Knowledge: The Biological Roots of Human Understanding, Boston, MA: Shambhala Publications

Maturana, H. and Varela, F., 1994. De máquinas y seres vivos – Autopoiesis: la organización de lo vivo, Editorial Universitaria, Santiago de Chile

Mayr, E. (1969). Footnotes on the philosophy of biology. Philosophy of Science, 36(2), 197-202.

Mayr, E. (1982). The place of biology in the sciences and its conceptual structure. In E. Mayr (Ed.), The Growth of Biological Thought. Diversity, Evolution, and Inheritance (pp. 21-82). Cambridge, Mass: Harvard University Press.

Mayr, E. (1991). The Ideological Resistance to Darwin's Theory of Natural Selection. Proceedings of the American Philosophical Society, 135(2), 123-139.

Mayr, E. (1991). The Ideological Resistance to Darwin's Theory of Natural Selection. Proceedings of the American Philosophical Society, 135(2), 123-139.

Mayr, E. (2001). The Philosophical Foundations of Darwinism. Proceedings of the American Philosophical Society, 145(4), 488-495.

Mayr, E. (2004). The Autonomy of Biology. Ludus Vitalis, XII(21), 15-27.

McNamara, P. (1999). Mind and Variability: Mental Darwinism, Memory, and Self (Human Evolution, Behavior, and Intelligence). Westport, CT.: Praeger.

Mitchell, M., & Taylor, C. E. (1999). Evolutionary Computation: An Overview. Annual Review of Ecology and Systematics, 30, 593-616.

Moray, N. (1999). Mental Models in Theory and Practice. In D. Gopher & A. Koriat (Eds.), Attention and Performance XVII: Cognitive Regulation of Performance: Interaction of Theory and Application (pp. 223–258): MIT Press.

Morecroft, J. and Sterman, J. 1994. *Modeling for learning organizations*, Productivity Press

Morin. E. (1991). La méthode – 4. les idées: leur habitat, leur vie, leurs moers, leur organization. Seuil

Nanay, B. (2002). The Return of the Replicator: What is Philosophically Significant in a General Account of Replication and Selection? Biology and Philosophy, 17, 109-121.

Nanay, B. (2009). Replication without replicators. Synthese, in press.

Nelson, R. R. (2007). Universal Darwinism and evolutionary social science. Biology and Philosophy, 22, 73-94.

Olaya, C. (2007). Can The Whole Be More Than The Computation Of The Parts? A Reflection On Emergence. In C. Gershenson, D. Aerts & B. Edmonds (Eds.), Worldviews, Science and Us. Philosophy and Complexity: Essays on Epistemology, Evolution, and Emergence (Vol. 81-98). London: World Scientific.

Olaya, C. (2008). The End of Control. International Journal Of Applied Systemic Studies, 2(1/2), 127-145.

Perkinson, H. J. (1984). Learning from Our Mistakes: A Reinterpretation of Twentieth-Centuty Educational Theory. London: Greenwood Press. .

Piattelli-Palmarini, M. (1989). Evolution, selection and cognition: from "learning" to parameter setting in biology and in the study of Language. Cognition, 31, 1-44.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. Behavioral and Brain Sciences, 13, 707-784.

Popper, K. (1972). Objective Knowledge. An Evolutionary Approach. Oxford, UK: Oxford University Press.

Polanyi, M., *The Tacit Dimension*, Peter Smith, 1983

Redorbit (2010): Mission Control During Apollo 13 Crisis. (http://www.redorbit.com/images/gallery/apollo_xiii/mission_control_during_apollo_13_crisis/54/24/index.html; 1/31/2011)

Reeke Jr., G. N., & Edelman, G. M. (1988). Real Brains and Artificial Intelligence. Daedalus, 117(1, Artificial Intelligence), 143-173.

Rescher, N. (2008). Process Philosophy. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy. Stanford, CA: The Metaphysics Research Lab, Stanford University.

Salas, D., & Olaya, C. (2009). No More "Learning": Selectionist Education for the Classroom. In Proceedings of the 21st Annual Meeting of the Human Behavior and Evolution Society. Fullerton, California State University, CA, USA.: HBES.

Salas, D., & Olaya, C. (2010). Darwin en el aula: conjetura sobre la posibilidad de una educación seleccionista. 12th International Conference on Philosophy of Education(Universidad de los Andes, Colombia).

Schaeken, W., & Johnson-Laird, P. N. (2000). Strategies in Temporal Reasoning. Thinking and Reasoning, 6(3), 193-219.

Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental Models and Temporal Reasoning. Cognition, 60, 205-234.

Schaffernicht, M. 2006. Detecting and monitoring change in models, System Dyamics Review 22(1):73-88

Schaffernicht, M. 2010a. Learning from rediscovering system dynamics models, Systèmes d'Information et Management 14(4)-2009, p. 87-105

Schaffernicht, M. 2010b. Causal loop diagrams as means to improve the understanding of dynamic problems: a critical analysis, Systems Research and Behavioral Science (Published online: 28 Jun 2010, DOI: 10.1002/SRES.1018)

Schaffernicht, M. and Groesser, S. A Comprehensive Method for Comparing Mental Models of Dynamic Systems, European Journal of Operational Research (ISI), doi:10.1016/j.ejor.2010.09.003;final version published online: 23-DEC-2010;

Schwaninger, M. and Groesser, S. (2008). System Dynamics as Model-Based Theory Building, Systems Research and Behavioral Science **25**:447-465

Seel, N. M., & Blumschein, P. (2009). Modeling and Simulation in Learning and Instruction: A Theoretical Perspective. In P. Blumschein, W. Hung, D. Jonassen & J. Strobel (Eds.), Model-Based Approaches to Learning (pp. 3-15). Rotterdam: Sense Publishers.

Simonton, D. K. (1999). Creativity as Blind Variation and Selective Retention: Is the Creative Process Darwinian? Psychological Inquiry, 10(4), 309-328.

Sipper, M. (1998). Fifty Years of Research on Self-Replication: An Overview. Artificial Life, 4, 237-257.

Smith, L. (2008). Reason Has Its Own Reasons. Human Development, 51, 148-152.

Smolin, L. (1992). Did the Universe Evolve? Classical and Quantum Gravity, 9, 173-191.

Smolin, L. (1997). The Life of the Cosmos. Oxford: Oxford University Press.

Sporns, O. (1994). Selectionist And Instructions Ideas in Neuroscience. International Review of Neurobiology, 37, 3-26.

Taylor, F. W. (1911). The Principles of Scientific Management. New York, NY: Harper & Brothers.

Vandierendonck, A., & De Vooght, G. (1996). Evidence for Mental-model-based Reasoning: A Comparison of Reasoning with Time and Space Concepts. Thinking and Reasoning, 2(4), 249-272.

Varela F. 1979. Principles of Biological Autonomy. Elsevier: New York.

Varela F. 1981a. Describing the logic of the living. In Autopoiesis: A Theory of Living Organization, ZelenyM (ed.). Elsevier: New York; Ch. 3.

Varela, F., 1981b. Le cercle creatif, in Watzlawick,, pp.329-345

Varela, F. J. (1993). When is a map cognitive? In E. Lazslo (Ed.), The Evolution of Cognitive Maps (pp. 99-104). Amsterdam: Overseas Publishers Association.

Varela, F. J. (1997). Patterns of Life: Intertwining Identity and Cognition. Brain And Cognition, 34, 72-87.

Varela, F. J., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The Organization of Living Systems, Its Characterization and a Model. Biosystems, 5, 187-196.

Varela, F. (1992). Whence perceptual meaning? A cartogtaphy of current ideas, in: Understanding origins – contemporary views on the origin of life, mind and society. Edited by Francisco Varela and Jean-Pierre Dupuy, Kluwer Academic Publishers, Dordrecht, Boston, London

Varela, F., Maturana, H. and Uribe, R., Autopoiesis, the organization of living systems, its characterization and a model, BioSystems 5 (1974): 187-196

Varela, F.J. (1995), "The Re-Enchantment of the Concrete. Some biological ingredients for a nouvelle cognitive science", The artificial life route to artificial intelligence, Steels, L., & Brooks, R. (eds.), Hove, UK: Lawrence Erlbaum Associates, pp. 11-22

Watzlawick, P. Die erfundene Wirklichkeit – Wie wissen wir, was wir zu wissen glauben. Beiträge zum Konstruktivismus. Munich, Piper, 1981

Whitehead, A. N. (1978). Process and Reality (D.R. Griffin & D.W. Sherburne, corrected ed.). New York, NY: The Free Press.

Yang, C. D. (1999). A Selectionist Theory of Language Acquisition. In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (pp. 429-435). San Francisco, CA: Morgan Kaufmann Publishers.

Ziman, J. (2000). Technological Innovation as an Evolutionary Process. Cambridge, UK: Cambridge University Press.