Draft in the format of the pdf template

An Application of Category Theory to Cognition

by Andrée C. EHRESMANN

(Work in collaboration with Jean-Paul VANBREMEERSCH)

Abstract

This paper gives an outline of the model MENS for a neuro-cognitive system. Based on Category Theory. this 'dynamic' model describes the generation of an 'algebra of mental objects' through iterative binding of synchronous assemblies of neurons. It shows that the degeneracy of the neural code (Edelman, 1989) is the property allowing for the formation of mental objects and cognitive processes of increasing complexity order, up to consciousness.

1. Introduction

Despite the huge progresses of brain research in the last 20 years, we do not understand the brain's large-scale organizational principles allowing for the emergence of cognitive processes such as perception, memory, thought, consciousness. Can we hope to find common processes at the basis of cognition, leading to an integrative neuroscience and cognitive science as rigorous as physics?

Interesting mathematical models of a local nature have been developed, most often based on non-linear differential equations, dynamic systems, graph theory, stochastic processes or information theory; they concern particular processes in specialized brain areas with specific cells and neuromediators. As various brain areas are heterogeneous both anatomically and functionally, such models cannot be extended to other areas or processes.

However there is a common process in brain dynamics, already noted by Hebb in the forties (H ebb, 1949): it is the formation, persistence and intertwining of more or less complex and distributed neuronal assemblies, whose 'synchronous' activation is associated to specific mental processes. This association is not 1-1 due to the "*degeneracy* property of the neural code" emphasized by Edelman " more than one combination of neuronal groups can yield a particular output…" (Edelman, 1989, p. 50). Thus the mental representation of a stimulus should be the common 'binding' of the more or less different neural patterns which it synchronously activates either simultaneously or at different times.

This binding process plays a central role in the mathematical model MENS (for Memory Evolutive Neural Systems) of a neuro-cognitive system of which we present an outline in this paper. This model proposes a common frame accounting for the functioning of the neural and cognitive system at different (micro, meso, macro) levels of description and across different timescales. It describes how different brain areas interact as hybrid systems to generate an "algebra of mental objects" (in the terms of Changeux, 1983) through iterative binding of more and more complex neural patterns, and how it leads to the emergence of flexible higher mental and cognitive processes.

MENS does not constitute a logic model of the invariant structure of the neuro-cognitive system; it is intended to give a 'dynamic model' sizing up the system 'in the making', with the variation over time of its configuration and successive specifications of its logic, It is based on a particular domain of mathematics, category theory. It will not help as such in explaining the neural mechanisms (such as Hebb rule) by which synchronous assemblies of neurons are formed, but it gives a frame for treating mental objects as 'higher order' neurons (called *category-neurons*) and for explaining the emergence of cognitive processes of increasing complexity.

2. Category Theory

Category theory, introduced by Eilenberg and Mac Lane (1945) in the early forties, has a unique status, at the border between mathematics, logic, and meta-mathematics. It was introduced to relate algebraic and topological constructs. In the late fifties, its foundational role in mathematics was made apparent, in particular through the introduction of adjoint functors and (co)limits by Kan (1957). Later its role in logic was emphasized by several authors: for example, in the theory of *topos* developed by Lawvere and Tierney, and in the *sketch theory* developed by C. Ehresmann. It makes a general concept of structure possible, and indeed it has been described as mathematical structuralism, providing a single setting unifying many domains of mathematics.

Category theory tries to uncover and classify the main operations of the "working mathematician"; for instance defining a general notion of sub-structure, of quotient structure, of product,... valid as well for sets, groups, rings, topological spaces... Now mathematical activity reflects some of the main operations that man does for making sense of his world: distinguishing objects (a tree, a fruit,...); formation, dissolution, comparison, and combination of relations between objects (the fruit is linked to the tree, these fruits have the same colour, one fruit is larger than another,...); synthesis of complex objects from more elementary ones (binding process) leading to the formation of hierarchies (complexification process); optimization processes (universal problems); classification of objects into invariance classes (formation of concepts)..

As all these operations are at the root of our mental life, and also of science, it explains that category theory begins to be applied in other scientific domains, in particular computer science, in the foundations of physics for studying quantum field theories, in biology (Robert Rosen, 1958) whose work is pursued by different authors.

Since 25 years, we have developed the theory of Memory Evolutive Systems (Ehresmann & Vanbremeersch, 1987, 2007), a model, based on a 'dynamic' category theory (incorporating Time and durations) for studying autonomous hierarchical natural systems such as biological systems, cognitive systems or social systems. Indeed, contrary to classical models which are adapted to study local problems, category theory allows taking account of different levels and covering at the same time the local, global, evolutionary and temporal aspects.

In particular it provides tools for studying the *binding problem*, including its possible *degeneracy*, leading to the emergence, through iterated *complexification* processes, of increasingly complex objects. It also explains how the *multi-scale self-organization* of the system can be generated internally, through the competition/cooperation between a net of internal functional "coregulator" sub-systems, each operating at its own rhythm and its own complexity level.

2.1. *Graphs. The graph of neurons*

Graphs have been used to represent networks of any nature: cellular networks, social networks, internet,.... Here we define a *graph* G as a set G_0 of objects A, B,…, called its *vertices*, and a set of oriented edges (or *arrows*) between them; an edge *f* from A to B is represented by an arrow $f\colon A \to B$. It is possible to have several arrows with the same source A and the same target B, and even 'closed' arrows (the source and target are identical). Let us remark that the term 'graph' is often restricted to the case where there is at most on arrow from a vertex to another, in which case the graph can be represented by a matrix.

A *path* of the graph from A to B is a sequence of consecutive arrows

$$(f_1, f_2, ..., f_n) \text{ with } f_1\text{: A} \rightarrow \text{A}_1 \,.\, f_2\text{: A}_1 \rightarrow \text{A}_2\,, \ldots , f_n\text{: A}_{n-1} \rightarrow \text{B}.$$

The paths of G form the *graph of paths of* G, denoted P(G): it has the same vertices as G but its arrows from A to B are the paths of G from A to B. We identify G with a sub-graph of P(G) by identifying an arrow f to the path (f) with f as its unique arrow.

If G and G' are two graphs, a *homomorphism p* from G to G' associates to each vertex A of G a vertex p(A) of G', and to each arrow f from A to B an arrow $p(f)$ from p(A) to p(B).

Example: The graph of neurons at an instant t: Its vertices, briefly called *neurons*, model the states N(t) of the neurons N existing at t (measured by their activity around t). An arrow from N(t) to N'(t) models a synapse f from N to N', labeled by its *propagation delay* around t and by its *strength* to transmit an activation from N to N'. The strength (negative if the synapse is inhibitory) varies according to *Hebb rule*: it increases if the activations of N and N' are correlated. The graph of paths of this graph will be at the root of our model; the *propagation delay* of a path is defined as the sum of those of its factors; and its *strength* the product of their strengths.

Graphs and their paths are not sufficient to account for the fact that several paths from N to N' may play equivalent functional roles, for instance different synaptic paths may transmit the same activation from N to N'. The notion of a category enriches that of graph by allowing a comparison of paths to distinguish "functionally equivalent paths".

2.2. *Categories*

A *category* is a graph equipped with an internal composition associating to a 2-path

$$(f, g) \quad \text{where } f\text{: A} \rightarrow \text{B and } g\text{: B} \rightarrow \text{C},$$

an arrow fg; A → C called their *composite.* The vertices are also called objects of the category and the arrows *morphisms*, or more simply *links.*

The composition satisfies 2 conditions:

(i) it is associative: for a path (f, g, h) we have $f(gh) = (fg)h$. It follows that each path, say (f_1, f_2, …, f_n) (whatever its length) has a unique composite (whatever its 2-2 decomposition) denoted $f_1f_2\ldots f_n$.

(ii) for each object A there is an arrow id_A from A to A, the *identity* of A, whose composite with any arrow of source or target A is identical to f.

A *functor p* from a category C to a category C' is a homomorphism of graphs which respects the composition and the identities, so that

$$p(fg) = p(f)p(g) \quad \text{and } p(\text{id}_\text{A}) = \text{id}_{p(\text{A})}.$$

The graph of paths of any graph G becomes a category when we define the composite of two consecutive paths as their convolution. The identity of an object A is the 'void' path from A to A. Each category is a quotient of the category of its own paths by the equivalence relation on paths: "two paths are equivalent if they have the same composite".

2.3. *The Evolutive System of Neurons*

If we take for graph G the graph of neurons at an instant t, the category of its paths is called the *category of neurons at t*, denoted NEUR_t. Its objects model the states of the neurons existing at t, the morphisms (also called *links*) are the synaptic paths between them, with their propagation delay and strength.

The category of neurons varies over time: some neurons 'die', new neurons are formed. The change from *t* to a later time *t'* is modeled by a functor, called *transition*, from a sub-category of $NEUR_t$ to $NEUR_{t'}$; it associates to the state at *t* of a neuron N its new state at *t'* provided that N still exists at *t'*, and similarly for the links.

The categories $Neur_t$ and the transitions between them during the life of the animal form the *Evolutive System of neurons* NEUR, whose components model the neurons. NEUR gives a description of the neural system at successive times; it does not explain how its dynamic is internally generated.

Definition. An *Evolutive System* K consists of:

(i) The time-scale T of the system and, for each *t* of T, a category K_t called *configuration* of the system at *t.*

(ii) For each time $t' > t$, a functor *transition* from *t* to *t'* from a subcategory of K_t to $K_{t'}$. These functors satisfying a transitivity condition.

A *component* of K is a maximal family of objects in the categories K_t related by the transition functors, so that the objects of K_t represent the successive states of the components. The configurations correspond to successive snapshots of the system, and the transitions describe what has changed between them, but not how the change has been produced.

3. **The binding process: Category-neurons and their links**

The binding process is ubiquitous in evolution: how do simple objects bind together to form a complex object forming "a whole that is greater than the sum of its parts "? for instance a wall compared to the heap of bricks of which it is formed. What are the different patterns of "parts" leading to the same whole? And what are the simple and complex interactions arising between such complex objects?

In Neuroscience, this problem has been emphasized by von der Malsburg and Bienenstock (1986), and we are going to show how it can be analyzed by using the categorical notion of *colimit* (or 'inductive limit, Kan, 1957). We know that a mental object (*e.g.*, the mental image of a simple stimulus) activates a synchronous assembly of neurons; but how such assemblies interact? The mental object can be represented by a neuron if there is a neuron N 'binding' the assembly, in the sense that N and the assembly as such have the same functional activating role. For instance Hubel and Wiesel (1962) has shown the existence of neurons representing a segment or an angle; and there are also neurons representing more complex but very familiar objects.

However generally there is no such "grand-mother neuron" (Barlow, 1972) binding the neural assembly activated by a mental object, and the same mental object can even activate different assemblies (due to the degeneracy property of the neural code, Edelman 1989). How to model the mental object as such and to determine how it interacts with neurons and other mental objects?

To answer this question, we enlarge NEUR into the *Memory Evolutive Neural System* MENS (Ehresmann and Vanbremeersch, 2007) whose components are conceptual objects, called *category-neurons* (abbreviated in *cat-neurons*) which model more and more complex mental objects and processes, The idea of this 'complexification process' is that a cat-neuron 'binds' (= becomes the colimit in MENS of) a synchronous assembly of (cat-)neurons, but binds also any other assembly of (cat-)neurons activated by the same mental object. The complexification process also indicates the 'good' links between cat-neurons. Thus cat-neurons can be thought of as virtual more complex neurons, and we can speak of assemblies of cat-neurons and iterate the construction. Successive iterations lead to higher order cat-neurons representing more and more complex mental objects.
First we describe the category $MENS_t$ whose objects model the states at *t* of existing or new mental objects.

3.1. *Patterns and their binding (as a colimit)*

An assembly of neurons synchronously activated at a time t (say by a stimulus S) is modeled by a pattern P in the category of neurons $NEUR_t$. A *pattern* is a family of neurons P_i interconnected by some distinguished links (= synaptic paths) f from P_i to P_j through which they transmit their activation to each other.

Such an assembly can activate a (cat-)neuron N if there are links from the different P_i to N which all transmit an activation of P_{ii} to N at the same time, taking account of the distinguished links of P. This operation cannot be represented in a simple graph. In a category it is modeled by a *collective link* from the pattern P to N.

Definition. A *collective link* from a pattern P to an object N consists of a family (s_i) of links s_i from P_i to N which satisfies the following equations:

$$s_i = fs_j \text{ for each distinguished link } f \text{ from } P_i \text{ to } P_j;$$

In particular this equation implies that the propagation delay of s_i is the sum of those of f and of s_j. so that P acts as a *polychromous* pattern (in the sense of Izhikevich & *al*,. 2004).

If the stimulus S is repeated or persists, the distinguished links of the pattern P activated by S are strengthened (via Hebb rule), and P takes an identity of own to act as a synchronous assembly. The long-term memory of S will be recorded in the category of mental objects $MENS_t$ by a cat-neuron M 'binding' P, so that it has the same functional role than P operating collectively (cf. Figure 1). For instance S could be a rectangle and P the pattern consisting in (the neurons activated by) its sides and vertices, with distinguished links from a vertex to the sides containing it. Then M would 'be' the mental image of S.

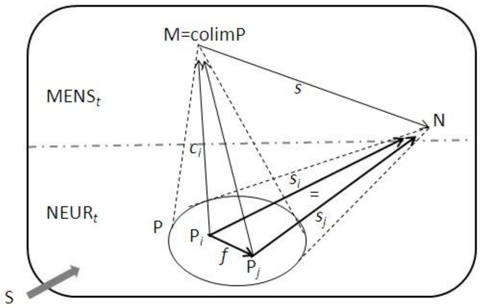


FIGURE 1. Collective link and colimit

Formally M becomes the *colimit* of P in $MENS_t$, meaning that the collective links (s_i) from P to any cat-neuron N are in 1-1 correspondence with the links s from M to N binding them. M is called (the state at t of) a *cat-neuron of level* 1 (to simplify, state at t is generally omitted); And P is called a *decomposition* of M.

Definition: In a category C, a pattern P is also called *diagram* and a collective link from P to N is called a *cone* with basis P and vertex N. We say that an object M *binds* P, or is a *colimit* (or *inductive limit) of* P if there is a cone (c_i) from P to M satisfying the *universal property*:

> Any cone (s_i) from P to N factors through a unique link s from M to N such that $s_i = c_i s$ for each i. We say that s *binds* (s_i).

3.2. *Category-neurons of level 1 and their links*

The degeneracy property of the neural code asserts that a stimulus S can activate more or less different synchronous assemblies of neurons, simultaneously or at different times, and these assemblies all have the same functional role since they correspond to the same mental image. In MENS, the cat-neuron M recording S must represent the invariant common to all these patterns; hence it must bind (= be the colimit of) each one of them. Though initially constructed (at t) to bind a particular pattern P, the cat-neuron M later takes its own identity as a component of MENS and can even disassociate from (the new state of) P at a later time t'. It is not a rigid memory (as in a computer), but a flexible multiform object which adapts to changing situations, so that S can later be recognized through the activation of M under different forms, even new forms not yet met (as long as the change is progressive enough).

What are the 'good links between cat-neurons in $MENS_t$? To describe them we need the following

Definitions (see Figure 2):

1. If P and P' are two patterns in a category C, a *cluster* from P to P' is a maximal set G of links between their components satisfying the following conditions:

(i) Each P_i has at least one link to a component of P'; and if there are several such links, they are correlated by a zigzag of distinguished links of P'.

(ii) The composite of a link in G with a distinguished link of P', or of a distinguished link of P with a link in G also belongs to G

2. If P and P' have colimits M and M' respectively, it follows from the universal property of a colimit that the cluster G 'binds' into a unique link cG from M to M', called a (P, P')-*simple link.*

3. Two patterns Q and P are *non-connected* if they have the same colimit M though there is no cluster between them binding into the identity of M. In this case M is called a *multiform* object, and the passage from Q to P a *switch* between decompositions of M.

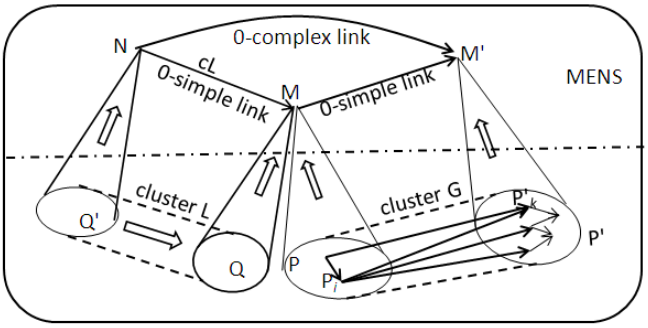


FIGURE 2. Cluster, simple and complex links

There are 2 kinds of links between cat-neurons (cf. Figure 2):.

(i) 0-*simple links.* Let M and M' be 2 cat-neurons of level 1, binding neural patterns P and P' respectively. A (P, P')-*simple link* from M to M', or 0-*simple link,* is a link binding a cluster G from P to P'. Such a link just translates at the level 1 of cat-neurons the fact that P can coherently activate the various components of P' through the links of G. A composite of 0-simple links binding adjacent clusters is 0-simple.

(ii) 0-*complex links.* The degeneracy property of the neural code has for consequence the existence of cat-neurons M which are *multiform*, meaning that M binds two non-connected neural patterns P and Q. In this case there are also 0-*complex links* which are the composites in the category $MENS_t$ of 0-simple links binding non-adjacent clusters separated by a switch, for instance (cf. Figure 2) a 0-complex link from N to M' composite of a (Q', Q)-simple link with a (P, P')-simple link, where P and Q are non-connected decompositions of M. They *emerge* at the cat-neuron level since they are not discernable directly through the neural components of N and M', though they depend on the 'global' properties of the neural level.

If N or M' is a neuron, the links from M to M' are similarly described by taking as decomposition for it the pattern reduced to this neuron.

3.3. *The hierarchy of cat-neurons*

Complex mental objects are constructed by binding together patterns of simpler ones. It will be the same for the cat-neurons representing them in MENS. Thus we define a hierarchy of cat-neurons which will be described iteratively, from the level 0 of neurons (also called *cat-neurons of level* 0) up.

Since we have described the links between cat-neurons of levels 0 (the synaptic paths) and 1 (the 0-simple and 0-complex links) in $MENS_t$, we can iterate the construction to obtain (the states at t of) cat-neurons of level 2 binding together patterns of cat-neurons of levels < 2, that is assemblies of neurons and of cat-neurons of level 1. The degeneracy property extends to the level 1, so that we obtain 1-simple and 1-complex links between cat-neurons of level 2, still constructed as above.

And by iteration of the process, we describe cat-neurons of increasing levels, with n-simple and n-complex links between them. The complex links reflect global properties of the lower levels which are not observable locally at these lower levels. They are at the root of the emergence of mental objects and processes of increasing complexity.

This completes the description of the category $MENS_t$: its objects are the (states at t of) cf cat-neurons of any level (included level 0 of neurons) representing mental objects, and the links of level n+1 are n-simple and n-complex links, for each n.

By construction, the cat-neurons are partitioned into levels, so that each cat-neuron M of level n+1 binds at least one pattern P with values in the full sub-category whose objects are of level $\leq n$; and some M are multiform. Thus, according to the following definitions, $MENS_t$ is a *hierarchical category* which satisfies the Multiplicity Principle.

Definitions. 1. A category is *hierarchical* if its objects are partitioned into levels such that each object of level n+1 is the colimit of at least one pattern contained in the levels < n+1.

2. An object M of level n+1 is *multiform* if it is the colimit of 2 lower levels patterns P and P' such that there is no cluster between them binding into the identity of M. The category satisfies the *Multiplicity Principle* if it admits multiform objects.

4. **The hierarchical Evolutive System MENS**

We have said that MENS is an Evolutive System. We have just described its configuration category $MENS_t$ at each instant t. he transition from t to t' maps the state of a cat-neuron existing at t to its state at t' if it still exists. Since the categories $MENS_t$ are hierarchical, we say that MENS is a *Hierarchical Evolutive System.*

Once formed at a time t, a cat-neuron M preserves its identity up to its 'death' (by decomposition or loss of enough components), though its decompositions can vary more or less quickly over time. We define the *stability span* of M at an instant t as the longest period during which M admits a decomposition P at t which remains a decomposition of M. This span is important for it imposes material constraints on the functioning of the system (cf. Section 5).

4.1. *The complexification process*

The categories $MENS_t$ model the configuration of the neural, mental and cognitive system at a time *t*. It changes over time by formation of new neurons, of new more or less complex mental objects and of links between them, but also of loss of some cat-neurons. The transition from *t* to *t'* must reflect these changes.

In other words, the transition corresponds to the realization of a procedure Pr with objectives of the following kinds: loss, decomposition or inhibition of some cat-neurons; formation, or preservation if it exists, of a new cat-neuron binding some pattern P' of already existing cat-neurons. (Cf. Figure 3.)

In categorical terms, this operation can be modeled as follows: the new configuration $MENS_{t'}$ at *t'* will be the *complexification* of $MENS_t$ with respect to the procedure Pr, and the transition is the partial functor from $MENS_t$ to $MENS_{t'}$.

We have explicitly constructed (EV 1987) this complexification. Its objects are the new states of the cat-neurons not suppressed by Pr, as well as new cat-neurons cP' binding the patterns P' specified by Pr (if 2 patterns to bind P' and P" have the same functional role, we take cP' = cP"). The links are simple and complex links defined as above.

The complexification process can be compared to the formation of specific words of a language from an alphabet: the letters being replaced by mental objects and the sequences of letters by any kind of pattern

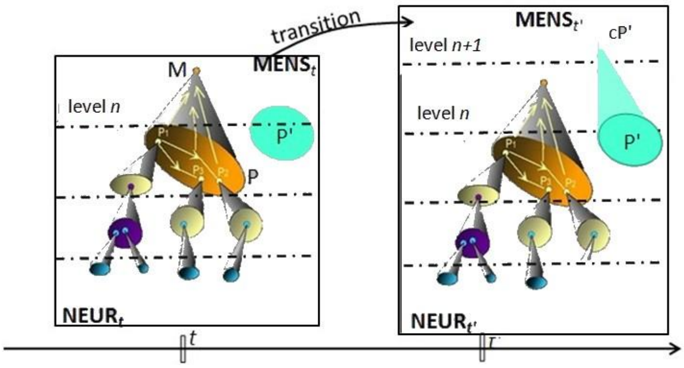


FIGURE 3. Hierarchical Evolutive System MENS

Remark. For higher animals, there will be another kind of cat-neurons obtained as the projective limit of some pattern Q. (A projective limit is a colimit in the opposite category obtained by inverting the direction of the arrows). When the procedure asks also for the formation of such 'classifying' cat-neurons, we speak of a *mixed complexification.* Its construction is more complicated; cf (Ehresmann and Vanbremeersch, 2007).

Definition. The (mixed) *complexification* of a category C for a procedure Pr is solution of the universal problem of constructing a category in which the objectives of the procedure are best realized.

Using the universal property of the complexification, we have shown (Ehresmann, 2009) that the propagation delays and strengths of the synapses can be extended to MENS, so that each link from M to M' has a *propagation delay* corresponding to the time necessary for transmitting an activation from M to M"; it also has a *strength* which satisfies an *extended Hebb rule*: If the activations of M and M' vary in the same direction, then the strength of the link increases.

4.2. *Complexity order of a mental object*

The *activation* (or recall) of the mental object represented by a cat-neuron M of level > 0 consists in the unfolding of one of its ramifications down to the neural level: first activation of one of its decompositions P into a synchronous assembly of cat-neurons of lower levels, then a decomposition of each component of P, and so on down to the *physical activation* of synchronous assemblies of neurons. At each step, there is a choice between various (possibly non-connected) decompositions, so that the activation of M has several freedom degrees leading to *multiple physical realizabilities* in hyper-assemblies (*i.e.* assemblies of assemblies of… assemblies) of neurons.

The ramifications of M have not all the same length. For instance a cube can be directly decomposed into its sides; or first decomposed into its faces, and then each face decomposed into its sides

We define the *complexity order* of M is the smallest length of a ramification. It is less or equal to the level of M. The level indicates the number of steps in which M has been constructed, while its order of complexity measures the smallest number of steps sufficient for its later activation.

The existence of cat-neurons of increasing complexity order is a main consequence of the degeneracy property of the neural code which, as we have shown, implies the existence of multiform cat-neurons. Indeed we have proved (Ehresmann and Vnbremeersch, 1996):

Theorem. *In a hierarchical category which does not satisfy the Multiplicity Principle, all the objects are of complexity order 1. If a category satisfies the Multiplicity Principle, so does a complexification, and then iterated complexifications may lead to the emergence of objects of strictly increasing complexity order.*

Corollary. *In* MENS, *there is emergence of cat-neurons of increasing complexity order, representing more and more complex mental objects or mental processes.*

5. MENS as a Memory Evolutive System

Up to now, MENS describes the successive configurations of the neuro-cognitive system and the structural changes between them reflected by the transitions. Now we are going to analyze how these changes are internally generated through the interactions of a variety of multi-scale local processes (each one possibly resorting to classical models based on differential equations).

The dynamic of the neuro-cognitive system depends on successive sensory, proprioceptive, motor, affective, cognitive… experiences which can be stored in a long-term memory and later recalled in analogue circumstances. Such experiences activate particular parts of the system and are processed by specific 'modules' or areas of the brain, from small specialized parts (the "treatment units" of Crick, 1994) such as the visual centers processing colour, to large areas such as the nuclei of the emotive brain (brain stem and limbic system) or the associative cortex.

5.1. *The Memory and the coregulators*

In MENS, the memory is modelled by an evolutive sub-system Mem which represents the memory of the system. A cat-neuron in Mem, called *record*, represents a mental object associated to an item (external object, signal, past event, internal state, sensory-motor or cognitive process,,..) which can be recognized and/or recalled through the activation of any of the ramifications of the record. Thanks to the multiplicity of its ramifications the record is a robust memory, though remaining flexible and constantly revised to account for changes and new situations.

Mem contains a sub-system Proc, the *Procedural Memory* in which the records, called *procedures* have links (or 'commands') toward the pattern of their effectors (*e.g.* motor commands of a specific movement), and a *Semantic Memory* Sem (cf. Section 6)..

To account for the modular organization of the brain, the dynamic of MENS is modulated by the cooperative/competitive interactions between functional sub-systems, called *coregulators,* which act as internal regulatory organs. A coregulator is based on a specific module of the brain, meaning that its cat-neurons have ramifications down to this module (so that they model hyper-assemblies of neurons of the module). The complexity order of a coregulator is the greatest order of its cat-neurons.

A coregulator CR operates stepwise at his own rhythm as a hybrid system, using its differential access to the memory Mem, in particular to recall the procedures specific of its function. Let us describe one step:

(i) At the beginning t of the step, the partial information accessible to CR is modelled by its *landscape* at t, acting as a working memory for CR. It is a category L_t whose objects are the clusters G from a cat-neuron B to CR (considered as a pattern) which have at least one link activating a cat-neuron in CR.

(ii) An adapted procedure Pr is selected through this landscape, using the differential access of CR to Mem to recall preceding analogue events. For instance if an object S is presented to a CR treating colours, the landscape will contain only information on the colour of S, and the sole objective of Pr will be to bind the pattern P of neurons activated by the colour.

(iii) The commands of the procedure are sent to its effectors in MENS. In the above example, the binding of P into a cat-neuron (called the CR-record of S) is realized by strengthening the distinguished links of P using Hebb rule. The dynamic by which the effectors realize their objectives during the continuous time of the step should be computable with the help of models based on differential equations implicating the propagation delays and strengths of the links between cat-neurons.

(iv) At the beginning of the next step, the result is evaluated (possibly through other coregulators) by comparing the anticipated landscape (which should be the complexification of L_t with respect to Pr) with the new landscape, and Pr and its result are recorded. We say that there is a *fracture* for CR is the objectives have not been attained.

5.2. *Interplay among local processes*

At each step of a coregulator CR, a procedure is selected on its landscape which only collects the partial information accessible to CR. However the commands of the procedure are not realized on the landscape, but through the commands they send to effectors of the system. At a given time the commands sent by the various coregulators should all be realized by the effectors. If there are few active coregulators, and their commands are not conflicting all will be realized.

The situation is different if there are several active coregulators, especially if they have higher order cat-neurons; since they have different functions and rhythms, the commands so sent to effectors on MENS can be conflicting. For instance to seize an object, the visual and motor commands should fit together. So there is need of an equilibration process to ensure the correlation of the different commands, possibly neglecting some of them. This process, called the *interplay among the coregulators*, leads to the *operative procedure* S° which will be 'really' implemented on the system.

The interplay searches for a best compromise between the more or less conflicting constraints, keeping as much as possible of the objectives of the various coregulators. In particular it takes benefit of the degrees of freedom of a multiform cat-neuron which can be activated through anyone of its lower level decompositions, with possible switches between them; the decompositions allowing for a better coordination are selected through a kind of Darwinian selection process. For instance, depending on the context, we can seize an object in the right or left hand.

In the interplay, an important role is played by *evaluating coregulators*, based on parts of the emotive brain which evaluate the procedures as a function of their consequences on the well-being of the animal.

The operative procedure S° actually carried out may by-pass the procedures of some coregulators thus causing dysfunction (temporary fracture or longer *dyschrony*) to them.

Among the causes of fractures, there is the non-respect of the *structural temporal constraints* of a coregulator CR imposed by the propagation delays and stability spans: for each *t*, we must have

$$p(t) << d(t) << z(t)$$

where $p(t)$ is the time-lag of CR (= mean propagation delay of the links in the landscape),, $d(t)$ is the *period of* CR (= mean length of the preceding steps), and $z(t)$ is the smallest stability span of the components intervening during the step.

The non-respect of these temporal constraints may lead to a backfiring of fractures between coregulators with heterogeneous complexity orders and rhythms.

Remark. Though the dynamic during one step of a given coregulator could be computed, the interplay does not seem open to 'classical' computations. In categorical terms we suppose it could resort to cohomological methods.

6. Semantic, higher cognitive processes, consciousness

The coregulators jointly participate in the development of the memory Mem over time, and in the formation of important functional sub-systems of Mem, the Semantic Memory and the Archetypal Core. Both will be essential for the emergence of higher cognitive processes, thought and consciousness.

The record of a stimulus S will result from the cooperation of the various coregulators which can access some attribute (colour, size,...) of S through their landscape. Each coregulator CR will have for objective to form its partial CR-record of S (cf. Section 5.1). The interplay among coregulators will bind together all the partial CR-records. And the recall of S can be done through the activation of these CR-records. This process can be compared to the learning and then recording of a tune by an orchestra, all the members cooperating in their specific way.

6.1. *The Semantic Memory*

For higher animals, a sub-system Sem of Mem, called the *Semantic Memory*, may gradually develop through the detection of specific invariants, which classify memorized items according to their main attributes; the invariance classes, called *concepts*, are modeled by classifying cat-neurons as sketched below. (For a more explicit construction of Sem, we refer to Ehresmann and Vanbremeersch, 2007.)

(i) Given a coregulator CR specialized for some attribute (e.g., color), we define the CR-*trace* of a record M as being the pattern activated in CR by a decomposition P of M. The CR-*concept* of M (or of the mental object it represents) is the projective limit C_M of the CR-trace of M; it is added to MENS through a mixed complexification process. Then C_M is the reflection of M in the evolutive sub-system $\mathrm{Sem}_{\mathbf{CR}}$ of Mem formed by the CR-concepts**.** Any record M' which admits C for reflection in $\mathrm{Sem}_{\mathrm{CR}}$ is called an *instance* of C**.**

(ii) More abstract concepts are obtained by iteratively adding colimits or projective limits of patterns of CR_i-concepts for various coregulators CR_i through mixed complexification processes.

The activation of a concept can be done through any of its instances with possible shifts between instances. Thus the interplay among coregulators acquires new freedom degrees when some of the intervening cat-neurons are concepts of records of high complexity order. Indeed, they can be activated through the activation of any of their instances, with possible shifts between these instance; and then through the unfolding of any ramification of an instance, with possibility of switches between its decompositions at the various levels.

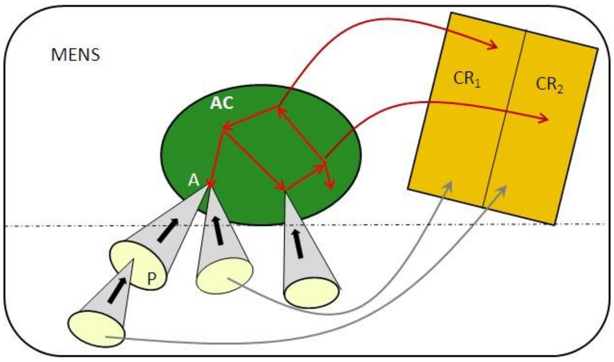
6.2. *The Archetypal Core*

The semantic memory will play a role in the development, from birth on, of a higher order sub-system AC of Mem which we call the *Archetypal Core* 'EV 1999).

AC is essential for the emergence of higher cognitive processes, in which a large number of coregulators of higher orders must collaborate. The interplay among them will necessitate collecting a large number of information, and analyzing them and operating on them will take some time. AC will act as an internal model of the Self which provides a stage and a motor for these operations.

AC is based on an integrative part of the brain called the *structural core* by Hagmann & *al.* (2008) who write: "This complex of densely connected regions in posterior medial cortex is both spatially and topologically central within the brain. Its anatomical correspondence with regions of high metabolic activity and with some elements of the human default network suggests that the core may be an important structural basis for shaping large-scale brain dynamics. "

AC consists of higher order cat-neurons, with ramifications down to the structural core, which integrate and intertwine recurring memories and experiences of different modalities, as well as notable events; they are connected by complex links which become stronger and faster (thanks to Hebb rule) along time. These links form *archetypal loops* which propagate very quickly the activation of an archetypal record back to itself, thus maintaining it. The activation resonates to lower levels via the unfolding of ramifications and switch between them, so that the activated domain of MENS is increased**.**

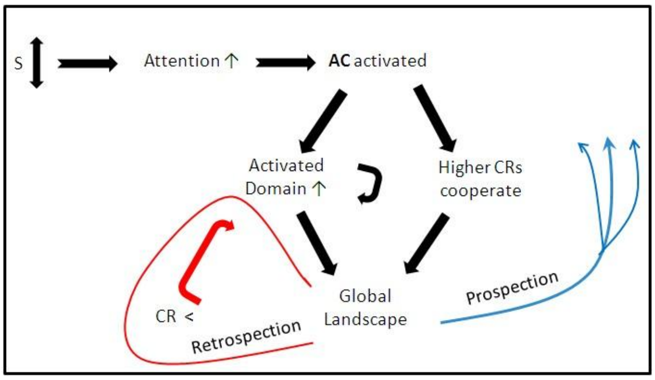


6.3. *Conscious processes*

We say that a coregulator is *conscious* if it is a higher coregulator, based on associative brain areas (compare to the "conscious units" of Crick, 1994), which is directly linked to some archetypal cat-neurons. Such coregulators will develop a process which allows an internal trespassing of temporal constraints thanks to a dynamic control of the memory, leading to more adapted responses; this agrees with ideas of Changeux (1983) and Edelman (1989) on consciousness.

An arousing situation or a non-expected event S (such as a fracture in a conscious coregulator), increases awareness, (through the activation of the reticular formation), that leads to the activation of some archetypal cat-neurons. It triggers, through archetypal loops, a self-maintained recollection of a large part of AC, and, by resonance to lower levels through ramifications (as said above), it activates a large domain of MENS. This activation is transmitted back to conscious coregulators, which can cooperate (through AC) to construct a *global landscape GL* uniting and extending their landscapes. *GL* is a mirror not of the present state but of the various evanescent traces recently accumulated in the working memory. It can be compared to Baars' "theater" (Baars, 1997).

The formation of GL gives a frame for thought and allows for the development of conscious processes presented as an integration of the time dimension through 2 possibly intermingled processes:



(i) A *Retrospection process* (toward the past) proceeds by abduction (in the sense of Pierce, 1903) to recollect information back in time thanks to the renewal of its activation via AC, so that it becomes observable in *GL*. Through several steps, it allows for analyzing the triggering event S and its possible causes, thus "sensemaking" of the present.

(ii) A *Prospection process* (toward the future) is then developed in the long term *GL*. Still using the motor role of AC, it iteratively constructs virtual landscapes in which sequences of procedures are tried with evaluation of their risk of dysfunction (by use of the temporal constraints and results of former similar events). Thus various "scenarios" are built. Once a scenario is selected, the retrospection process allows back-casting to find sequences of procedures (implicating coregulators of various levels) able to realize this long term program.

Conclusion

MENS proposes a "theory of mind", in which a hierarchy of mental objects and processes emerges from the functioning of the brain, through the iterative binding of neuronal assemblies. We show that *the degeneracy property of the neural code* is the characteristic which makes this emergence possible, and we explain how it allows the development of a flexible memory, with a central part, the Archetypal Core AC at the basis of the self and of the formation of higher cognitive processes up to consciousness is seen as the integration of past and future. The following table indicates the correspondences between Brain, MENS and Mind.

BRAIN	**MENS**	**MIND**
Neurons	Cat-neurons level 0	
Synchronous neural assemblies	Cat-neurons level 1	Simple mental objects
Synchronous hyper-assemblies of neurons	Cat-neurons level *n*	Algebra of mental objects
Degeneracy of the neural code	Multiform cat-neuron Complex links	Emerging mental objects
Structural core of the brain	Archetypal Core	Self
Consciousness loop	Global Landscape Retro-/ Pro-spection	Consciousness

References

Baars, B. J., 1997, *In the theatre of consciousness: The workspace of the mind*, Oxford University Press. Oxford.

Barlow, H.B., 1972, Single units and sensation: A neuron doctrine for perceptual psychology, *Perception* 1, 371-394

Changeux, J.-P., 1983, *L'homme neuronal*, Fayard, Paris.

Crick, F., 1994, *The Astonishing Hypothesis*, Macmillan Publishing Company, New York.

Edelman, G.M., 1989, *The remembered Present*, Basic Books, New York.

Eilenberg, S. and Mac Lane, S., 1945, General theory of natural equivalences, *Trans. Am. Math. Soc*. 58, 231-294.

Ehresmann, A. 2009, A propos des systèmes évolutifs à mémoire et du modèle MENS, Compte-rendu du Séminaire Intinérant des Catégories (Editeur R. Guitart), Paris, 2009,
http://www.math.jussieu.fr/~guitart/docpp/sic24octobre2009paris.pdf

Ehresmann, A.C. and Vanbremeersch J.-P., 1987, Hierarchical Evolutive Systems: A mathematical model for complex systems, *Bull. of Math. Bio*. 49 (1), 13-50.

Ehresmann, A.C. and Vanbremeersch J.-P., Multiplicity Principle and emergence in MES, *Journal of Systems Analysis, Modelling, Simulation* 26, 81-117.

Ehresmann, A.C. and Vanbremeersch J.-P., 2007, *Memory Evolutive Systems: Hierarchy, Emergence, Cognition*, Elsevier, Amsterdam.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Van J. Wedeen and Sporns, O., 2008, Mapping the Structural Core of Human Cerebral Cortex, *PLoS Biology* 6, Issue 7, 1479-1493. Online: www.plosbiology.org

Hebb, D. O., 1949, *The organization of behaviour*, Wiley, New York.

Hubel, D.H. and Wiesel, T.N., 1962, Receptive fields..., *J. Physio*. 160 (1), 106-154.

Izhikevich, E.M., Gally, J.A. and Edelman, G.J., 2004, Spike-timing Dynamics of Neuronal Groups, *Cerebral Cortex* 14, N 8, Oxford University Press.

Kan, D. M., 1958, Adjoint Functors, *Trans. Am. Math. Soc.* 89, 294-329.

Malsburg C. (von der) and Bienenstock E., 1986, Statistical coding and short-term synaptic plasticity, in *Disordered systems and biological organization*, NATO ASI Series 20, Springer, 247-272.

Pierce, C.S., 1903, Abduction and induction, in *Philosophical writings of Pierce* (Buchler, J., Ed.), Dover Publications, New York, 150-156.

Rosen, R., 1958, The representation of biological systems from the standpoint of the Theory of Categories, *Bull. Math. Biophys.* 20, 245-260.

Université de Picardie Jules Verne
Faculté des Sciences, Mathématiques
33 rue Saint-Leu, F-80039 AMIENS
France
ehres@u-picardie.fr
http://ehres.pagesperso-orange.fr