

# Toward a Cognitive System Algebra: A Perception/Action Perspective

P. Gaussier

Neuro-cybernetic team, Image and Signal processing Lab.,  
Cergy Pontoise University / ENSEA  
6 av du Ponceau, 95014 Cergy Pontoise, France  
email: gaussier @ensea.fr

**Abstract.** We propose a cognitive system algebra (CSA) useful to represent cognitive systems and to analyze them. This formalism is based on a deep structural link that should exist between Perception and Action in any autonomous system. The main idea is that Perception and Action are equivalent to voltage (tension) and current in electrical circuits. An energetic matricial representation  $\int Per.Ac^T dt$  is also introduced. In a simple case, maximizing the rank of this matrix allows to retrieve classical learning rules used to model conditioning mechanisms. We use it also to justify some simple algebraic simplifications rules in the case of a cognitive system in an equilibrium state (after learning stabilization).

## 1 Introduction

In recent brain modeling works, we assist to the presentation of more and more detailed but also complex architectures that become difficult to analyze and to compare. In the design of autonomous robots and animat systems, the problem is even more important since several schools of formalism exist: the classical IA formalism (Newell, 1982; Chatila, 1995), the Brook's subsumption (Brooks, 1986) and all the other behavior based approaches (Pfeifer and Scheier, 1996; Schöner et al., 1995). More and more researchers wonder about the validity and relevance of the current works since we do not know neither how to compare 2 cognitive systems (CS) nor to be sure an architecture cannot be simplified. Moreover, some open questions are: how to predict the major behavioral properties linked to a given architecture? How to connect adaptation and learning performances to the complexity of the architecture? How to measure a kind of energy consumption linked to the adaptiveness and the reactivity of the system (ability to show the simplest and most appropriated behavior in a given static or dynamic environment)?

Several interesting directions of research have been proposed in previous works that try to overstep the old theoretical frame of the cybernetics proposed by Wiener (Wiener, 1961) or Ashby (homeostat principle (Ashby, 1960)). In the animat community, the Steel's proposition about a mathematical way to analyze robot controllers was interesting but limited to pure behaviorist systems (Steels, 1994). Another study of Smithers showed the difficulty of characterizing the complexity (in terms of dynamical systems) of a simple obstacle avoidance behavior (difficulty to measure the fractal dimension of the phenomenon (Smithers, 1995)). Other interesting formalisms have been proposed in the frame of dynamical game theory (Ikegami, 1993). The searches in formal logic, signal processing, information theory, automatic and Neural Networks (N.N.) tackle some fundamental properties of CS. Each of them focuses on one particular feature but not on the global properties that make the system really "cognitive". For instance, the interpretation of neural dynamics in terms of dynamical systems (chaotic or not (Kelso, 1995; Schöner et al., 1995; Tsuda, 2001; Daucé et al., 1998; Kuniyoshi and Berthouze, 1998; Berthouze, 2000)) seems a very promising direction of research but also a way of

building and understanding large networks (Albus, 1991; Taylor, 1995).

The formalism, we propose, is based on our conviction that a deep structural link exists between Perception and Action (Berthoz, 1997; Pfeifer and Scheier, 1999; Gaussier and Zrehen, 1995) (this can be seen as another way to say that the cognition must be understood as a dynamical system). Specifically, we will emphasize that Perception and Action are the two faces of the same entity or coin. In any part of a CS, if an Action can be defined, this means there is an associated Perception and vice versa. We will suppose : *the perceptions and actions in a CS are equivalent to the voltage and current in an electrical circuit.*

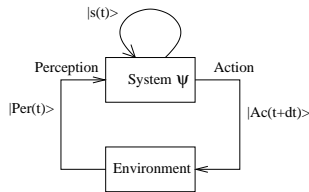


Figure 1: Typical system/environment interactions.

In the next sections, elementary operations (addition, composition...) will be defined and used to simplify a set of equations describing a CS. An energy measure will be introduced and used to compare the complexity of different architectures. We will show the LMS (Least Means Square) rule can be deduced from the minimization of an energy loss built from the (Perception, Action) couple. These energetical considerations will also be used to justify some algebraic simplification rules of the formal description of some simple CS (rules valid after learning stabilization). Finally, extensions and limitations of the proposed formalism will be discussed.

## 2 Basic formalization of a CS

We introduce here a mathematical formalism to manipulate CSs. The input and output of a CS will be represented by vectors in the “bra-ket” notation<sup>1</sup>. An input or output vector  $x$  (column vector) will be noted  $|x\rangle$  with  $|x\rangle \in R^{+m}$  while its transposed vector will be noted  $\langle x|$ . Hence  $\langle x|x\rangle$  is a scalar representing the square of  $|x\rangle$  norm. The multiplication of a vector  $|x\rangle$  by a matrix  $A$  is  $|y\rangle = A|x\rangle$  with  $|y\rangle \in R^{+n}$  for a matrix  $A$  of size  $n \times m$ .

<sup>1</sup>The formalism is inspired from Hilbert space used in quantum mechanics. Nevertheless, in our case it is not an Hilbertian space since the operator will not be linear...

A CS is supposed to be made of several elements or nodes or boxes associated with input information, intermediate processes and output (command of actions).

### Definition 1 “Cognitive System”

A CS  $\psi$  is a function associated to a control structure that belongs to the space of the cognitive control structure  $C_s$  ( $\psi \in C_s$ ). We postulate  $C_s$  associated operators that will be defined now is an algebra.  $0$  denotes a null CS (no output whatever the inputs are).

If  $|s(t)\rangle$  represents the current internal state of a CS,  $\psi$  allows to compute the next action and the future internal state of the CS:

$$(|Ac(t+dt)\rangle, |s(t+dt)\rangle) = \psi(|Per(t)\rangle, |s(t)\rangle) \quad (1)$$

To build a CS from the sum of 2 cognitive systems, we have to analyze 2 different cases. In the case both CS have completely different input and output, the problem is trivial. We will consider that the 2 systems have the whole set of inputs and outputs with null connections between the not used input/output (see fig. 2). In the more frequent case where two systems have some common inputs/outputs, the problem is to merge the outputs (there is no problem with the inputs because it is the same thing to consider 2 identical input groups or a single one with two times more connections). In the sequel, we will note  $O(\psi)$  the vector representing the concatenation of all the output groups of the CS  $\psi$ .

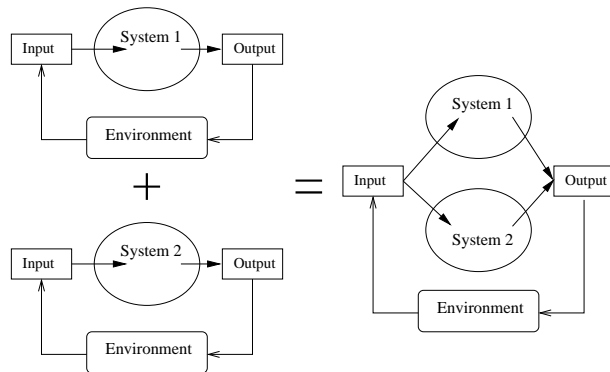


Figure 2: Merging of 2 cognitives systems. The sum of 2 CSs is also a CS.

If both systems propose two different outputs a problem can arise to merge their outputs. Indeed, if two different actions are proposed, an average action can be triggered or one of them can be chosen.

The main constraint in an autonomous system will be the temporal stability of the decision making. It would be stupid to choose an action and in the next iteration to change it to the opposite action. The formalism of the neural fields theory is a very good solution to solve all these problems (Amari, 1977; Schöner et al., 1995). So we will suppose that all the output groups of a CS are competitive structure controlled by neural field equations so that different outputs of the same modality can be merged easily.

We summarize now some basic properties used in our Cognitive System Algebra (CSA):

**Property 1 Addition**

*The sum of 2 cognitive systems is a CS (the addition is an internal law of  $C_s$ ). For any couple  $(a, b) \in C_s \times C_s$ , there exists a unique  $c \in C_s$  such as  $a + b = c$ . The addition is commutative and associative. Let take  $a, b$  and  $c \in C_s$  then  $(a+b)+c = a+(b+c)$ . 0 is the neutral element for the addition. If  $a \in C_s$  then  $a + 0 = 0 + a = a$ .*

**Property 2 Vectorial product**

*The product of 2 cognitives systems  $\psi_1$  and  $\psi_2$  is also a cognitive system  $\psi_3 = \psi_1 \otimes \psi_2$  corresponding to the vectorial product of the output of  $\psi_1$  and  $\psi_2$ . The resulting system is the sum of the 2 CS plus a new group build from the output of  $\psi_1 \cdot \psi_2^T$*

$$\psi_1 \otimes \psi_2 = \psi_1 + \psi_2 + O(\psi_1) \cdot O^T(\psi_2)$$

**Property 3 Product by a scalar**

*The product of a cognitive system  $\psi$  by a non null scalar constant does not change the cognitive properties of the resulting system  $k\psi = \psi$ . 0 is an absorbent element:  $0 \cdot \psi = 0$ . 0 will denote an integer value or a CS as well.*

$$\forall \psi \in C_s \text{ and } k \in N^*, \quad \sum_{i=1}^k \psi = k \cdot \psi$$

**Property 4 Composition**

*Two CS can be composed (push-pully connected) to create a new CS.*

$$\psi_1 \circ \psi_2(|X \rangle) = \psi_1(\psi_2(|X \rangle))$$

*with Id the neutral element. If  $f$  and  $g \in C_s$  then  $f \circ g \in C_s$*

This list of properties is not exhaustive but it shows the possibility to manipulate CS via algebraic rules. A complete mathematical study is out of the scope of the present paper.

### 3 Flow and effort in a CS

The bond graph theory is a very interesting framework to analyze complex systems and to simulate them (Rosenberg and Karnopp, 1983). It has been proposed as a general tool to describe any physical, biological or chemical system composed of sub systems that exchange energy between them. It has been shown that any physical system can be expressed in terms of effort and flow equivalently to voltage and current in electrical circuits for instance. After writing the physical system under this formalism, it is possible to use general rules to simplify, analyze and simulate the system without any knowledge about the original problem.

The bond graph theory considers two kinds of junctions: serial and parallel. In the case of a serial connectivity, the flow is the same in all the branches and the total effort is the sum of all the efforts (see fig. 3). On the opposite, when subsystems are connected in parallel (derivation), the effort is the same in all the system, but the total flow is the sum of all the different flows that cross the subsystems (see fig. 4).



Figure 3: Serial connectivity. Typical representation of a functionalist architecture.

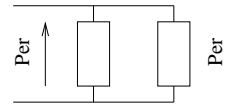


Figure 4: Parallel connectivity. Typical case of a behaviorist system in which all the behavioral blocks are supposed to be orthogonal to each others.

These 2 kinds of connectivity correspond to the 2 major approaches in cognitive sciences and robotics: functionalist and behaviorist approaches<sup>2</sup>. Indeed, fig. 3 corresponds to the classical decomposition of a problem in boxes used in sequence: preprocessing, classification, decision taking... On the contrary, the second architecture (fig. 4) is very similar to the Brook's subsumption (excepted that here nothing is said about the arbitration between the different proposition in case of contradiction (Brooks, 1986)).

<sup>2</sup>Note that if the present approach is valid the opposition between functionalist and behaviorist theories should vanish.

The interest of the electrical analogy does not only rely upon the way subparts of a CS can be combined to create a more complex CS. The analogy can also be applied to describe 3 fundamental basic behaviors of cognitive systems:

- **Resistive** element  $R$ . The action is proportional to the perception:  $|Per \rangle = R|Ac \rangle$  or  $|Ac \rangle = R^{-1}|Per \rangle$  (purely reactive system such as reflex systems or direct Perception/Action associations).
- **Capacitive** element  $C$ . The Action is proportional to the derivative of the Perception  $|Ac \rangle = C \frac{d|Per \rangle}{dt}$ . If the perception remains constant, the action vanishes (classical habituation mechanism in psychophysics and neurobiology).
- **Inductive** element  $L$ . The action is proportional to the perception integral  $|Ac \rangle = \frac{1}{L} \int |Per \rangle dt$ . Cyclic perception implies null action whereas a constant perception implies intensive action (sensitization mechanism).

In simple cases, the action vector  $|Ac \rangle$  can be expressed as the product between a perception vector  $|Per \rangle$  and a conductance matrix  $[G]$  such as:

$$|Ac \rangle = [G]|Per \rangle$$

One can notice that the majority of control architectures developed in the animat approach use only “resistive-like” elements. In the next section, we will show it is also possible to find the equivalent representation of a planning system in the case it receives a pure sinusoidal stimulation.

To sum up, we can consider that any element of a CS filters an input vector according to a weights matrix  $W$ . The final outputs are modified according to a non-linear function and a pattern of interactions between the elements of the same block (lateral interactions in the case of a competitive structure, recurrent feedback in the case of a dynamical system...). Hence, in the following we will speak about these elements as “neurons” even if they can be more complex algorithmic elements in other programming languages (any “if...then...else...” treatment can be expressed in terms of dynamical lateral interactions but the reciprocal is false). All the processes performed at the level of the neuron map are represented by a given operator,  $\mathbf{k}$  for instance. We will use the following formalism to represent all these processes:

$$|y \rangle = \mathbf{k}|W|x \rangle$$

The operator  $\mathbf{k}$  controls the way the weight matrix is combined with the input. It can be a product (classical neurons) or an explicit distance measure like in Kohonen maps:

$$\begin{aligned} |y \rangle &= \mathbf{k}|W|x \rangle & (2) \\ W &= (|w_1 \rangle, |w_2 \rangle, \dots, |w_n \rangle) \\ y_i &= \| |x \rangle - |w_i \rangle \| \end{aligned}$$

where  $\| |x \rangle \|$  is the norm of the vector  $|x \rangle$ .

The interest of this formalism is that we can also express all the simple vectorial transformations such as translation, rotation... of an input vector.

**Note 1** In the general case, the operator  $\mathbf{k}$  is non linear<sup>3</sup>. That implies  $\mathbf{k}$  is not distributive. Hence, in general:

$$k_1|W_1|(k_2|W_2|x \rangle + k_3|W_3|y \rangle) \neq k_1|W_1|k_2|W_2|x \rangle + k_1|W_1|k_3|W_3|y \rangle$$

When several groups are connected to the same group, we can write:

$$\begin{aligned} |y \rangle &= c|F_1|x_1 \rangle + \dots + c|F_i|x_i \rangle + \dots + c|F_n|x_n \rangle \\ \Rightarrow |y \rangle &= c \left( \sum_{i=1}^n F_i |x_i \rangle \right) \end{aligned}$$

It represents a right-side factorization of the operators on the same input vector.

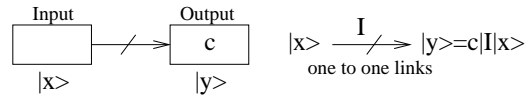


Figure 5: Unconditional “one to one” connections between two groups used as a reflex link. Left image is the graphical representation and right image is the formal notation.

In practice, we will distinguish 2 main types of connectivity between different parts of a CS. The first type is the “one to one” connection used for instance to define reflex mechanisms (see fig. 5) and represented by  $\mathbb{I}$  (identity connection).

<sup>3</sup>Indeed, in the case of a CS defined by eq. 1, the operator  $\mathbf{k}$  depends of the past.  $\mathbf{k}(t + dt) = f(\mathbf{k}(t), |Per(t) \rangle)$  where  $f$  is a function that can be defined from  $\psi$ .

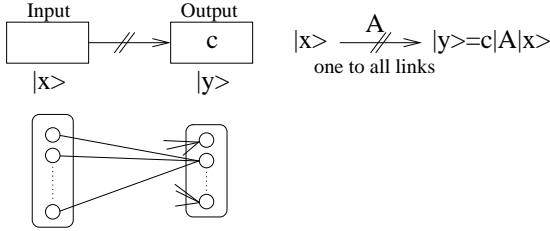


Figure 6: "One to all" connections with a competitive group representing the categorization of the input stimulus at the level of the output group. Left image is the graphical representation and right image is the formal notation.

The second type of connectivity concerns the "one to all" connections (see fig. 6) which are used for pattern matching processes, categorization... or all the other possible filtering. "One to all" connections will be represented in general by a  $A$  ( $A$  for all).

### 3.1 Analysis of a planning system

This analogy with electrical components may seem odd, so we will illustrate it with a "high level" cognitive problem. Let's consider a N.N. devoted to planning. We have shown, in previous works, that a group of neurons fully interconnected and using a MAX operator can be used as a cognitive map (Gaussier et al., 2000b; Quoy et al., 1999) (see fig. 7).

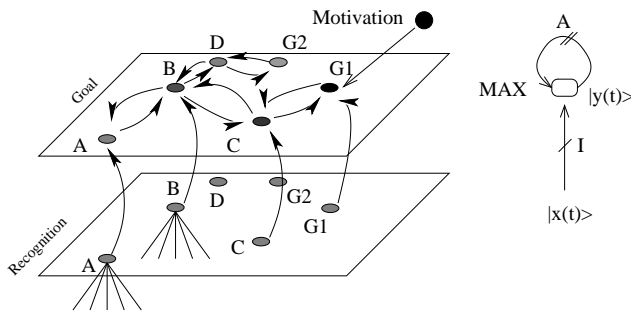


Figure 7: Simplified representation of a NN used to plan robot movements in a complex environment. The connexions between neurons on the upper map represent the map of the known environment.

A simple Hebbian learning rule, is sufficient to learn the connections between already discovered places. The MAX operator allows to back-propagate on the network a "motivation" (goal) so that the robot has only to follow the maximum gradient of nodes activation to find the shortest path to reach the goal. The vector  $|y \rangle$  representing

the activation of the set of planning neurons can be described as follows:

$$|y(t + dt) \rangle = \text{Max} | (A|y(t) \rangle + I|x(t) \rangle) | \quad (3)$$

with  $0 < A_{ij} < 1$  representing the learned connections on the cognitive map (links from place to place). The inputs (vector  $|x \rangle$ ) are connected to the planning neurons through "one to one" connections  $I$ .

Here, we consider such a map after learning and we stimulate one neuron with a sinusoidal signal (something that could be equivalent to activate more or less a given drive). The input vector  $|x \rangle$  can be written  $x_0 = 1 + \sin(\omega.t)$  and  $x_j = 0 \forall j \neq 0$ . Fig. 8 show the input signal and the signal recorded on 1 neuron on the output vector  $|y \rangle$  (note that the result should be almost the same with other kind of analogical path planning like the resistive grids for instance (Bugmann et al., 1995)).

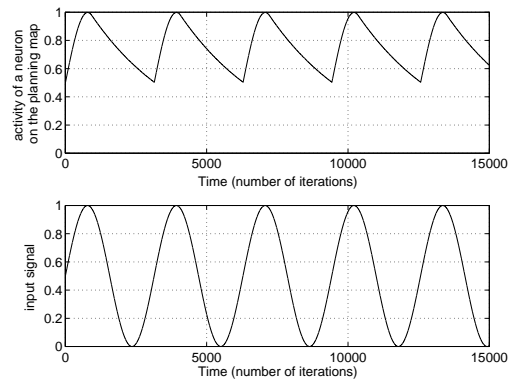


Figure 8: Activity of a neuron in a cognitive map (planning mechanism) according to a sinusoidal input stimulation imposed on a neuron (lower part) in the neighborhood of the recorded neuron (higher part - all synapses have the same value  $A_{ij} = 0.9997$ ).

The planning mechanism can be analyzed not only in terms of its algorithmic properties but also in terms of an equivalent electrical circuit or automatic device. If the input is constant (or its frequency is very high), the signal recorded on a neuron in the neighborhood of the stimulation is constant (or almost constant). When the period decreases, a capacitive effect can be perceived in addition to the obvious non linear behavior of the system. The phase of the output signal is shifted according to the input signal and its intensity tends to decrease less than the input. This phenomenon is linked to the value of the synaptic weights which is lower than 1 and is equivalent to a decay term

in the computation of the neuron activity (MAX operator).

Hence a cognitive map can be considered as an energy storage mechanism and its cognitive efficiency could be measured in terms of its capacity to store and deliver quickly the energy the system needs to face to a given situation.

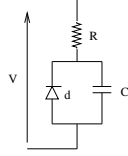


Figure 9: Equivalent electrical representation of a planning system excited by a periodical signal.

### 3.2 Basic simplification rules

Now, we can try to simplify some trivial networks (see fig. 10 and 11) to deduce basic simplification rules.

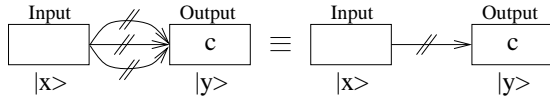


Figure 10: 2 groups linked with several one to all connections links (several times the same connections between 2 nodes).

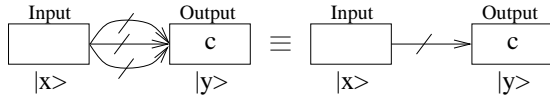


Figure 11: 2 groups linked with several one to one reflex connections links (several times the same connections between 2 nodes).

If we have a CS reduced to 2 reflex systems push pully connected (see fig. 12) then we can write  $|y \rangle = c|I|x \rangle$ ,  $|z \rangle = c|I|y \rangle$  (the input of CS2 is the output of CS1) and finally  $|z \rangle = c|I|c|I|x \rangle$  where  $c$  is the operator associated to a competitive group (i.e. WTA, Neural Field, ...).

In the case of a group of neurons of the same size and connected with one to one non modifiable links, it is trivial to prove that the output is directly a copy of the first input and that the intermediate groups can be suppressed (network simplification) because all the groups are equivalents<sup>4</sup>. This means

<sup>4</sup>It is possible to create a bijection between the activation of a given neuron in a first group and the activation of an-

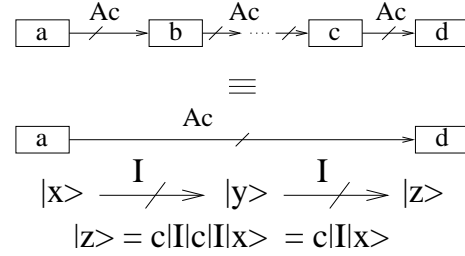


Figure 12: Simplification of a cascade of reflex mechanisms (unconditional connections)

the network fig. 12 is also equivalent to the network shown fig. 13.

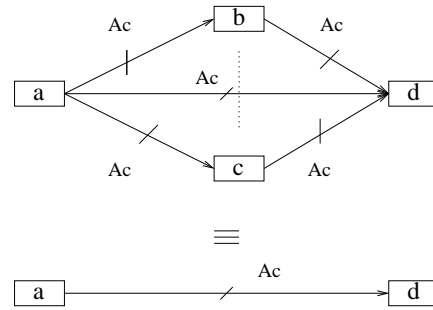


Figure 13: Equivalence of a cascade of unconditional connections between a set of group of neurons with a parallel structure of connections. Each group has the same kind and the links are "one to one" connections.

$$|y \rangle = c|I|c|I|x \rangle = c|I|x \rangle \text{ implies that}$$

$$\forall c \quad I|c|I \equiv I \quad (4)$$

but also

$$\forall c \quad c|I|c \equiv c \quad (5)$$

Note that 2 operators are equivalent if their exchange leave the vector output unchanged. In the case of a cascade of competitive structures (see fig. 12) that learn the output of the previous structure with one-to-all connections then the system can be simplified  $|y \rangle = c|A_1|c|A_2|x \rangle = c|A'|x \rangle$  implies:

$$\forall c \quad A_1|c|A_2 \equiv A' \text{ and also } \forall c \quad c|A|c \equiv c \quad (6)$$

In the case of derivative connections coming from and going to the same input/output group, simplification are possible. For instance, if we have  $|y \rangle =$  other neuron in a second group. Both sets of neurons can be considered as equivalents.

$c|(A|x \rangle + A|x \rangle)$  then  $|y \rangle = c|(A + A)|x \rangle = c|A|x \rangle$  which implies that

$$\forall F, c, \lambda \geq 1, \quad c|(\lambda F)|x \rangle \equiv c|F|x \rangle$$

is true only if  $c$  is a competitive structure. Let's notice that in the same case, we have also:

$$\forall F, c, \lambda \geq 1 \quad c|F|(\lambda|x \rangle) \equiv c|F|x \rangle$$

Now, we can use these rules to simplify a cascade of groups connected with one-to-one connections (see fig. 12 and fig. 13). We obtain the following equivalent relations:

$$\begin{aligned} |d \rangle &= c|\mathbb{I}|c'|\mathbb{I}|c''|\mathbb{I}|a \rangle \\ |d \rangle &= c|\mathbb{I}|a \rangle + c|\mathbb{I}|c'|\mathbb{I}|a \rangle + c|\mathbb{I}|c''|\mathbb{I}|a \rangle \\ |d \rangle &= c|\mathbb{I}(1 + c'|\mathbb{I} + c''|\mathbb{I})|a \rangle \\ |d \rangle &= c|\mathbb{I}|a \rangle \end{aligned}$$

The addition is commutative so a lot of other expressions are possible and a lot of simplification since

$$1 + x|W = 1$$

1 is an absorbent element for the addition whatever the operator  $x$  is.

A simple example of CS simplification is shown in appendix. Our first results suggest the proposed algebraic rules are efficient: the equivalent CS resulting from the CS algebraic simplifications seem to preserve the major functionalities of the departure system. If we come back to the electricity analogy, we must notice some subtle differences in order to obtain coherent results:

**Conjecture 1** *In the case of an adapted system, the current is null between the directly connected perception groups (they must have the same value). Hence a line of several perception groups can be simplified since they represent the same perception (null current, same tension).*

Note that the difference of perception between the different boxes must be null (the difference of perception measure should be defined up to a permutation). If the pattern of connectivity and the size of the groups and the interactions between neurons in the different maps are congruent to each others, the same is also true for the one to all learnable connections in the case of a "Winner Takes All" (WTA) or other self-organized structures.

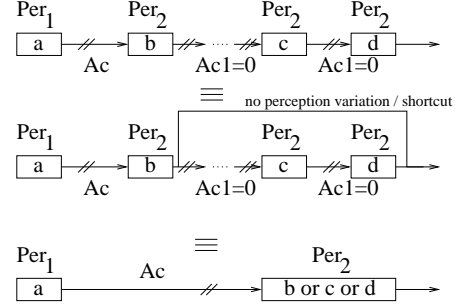


Figure 14: Simplification of a cascade of learnable connections between a set of group of neurons. Each group has the same kind and the links are "one to all" connections. If 2 groups are associated to the same Per then the intermediate group as a null impedance and the CS can be simplified.

Indeed, in fig. 14 after learning, at the level of the groups  $b, c$  and  $d$ , a unique and stable representation of the input ( $a$ ) is built. Hence, there exists an isomorphism of the activity between the different intermediate groups in the cascade. These intermediate groups can be suppressed without any change in the final output. Of course, if the connectivity is limited to a particular input neighborhood, several layers of neurons can be necessary before a complete coding of the global input, and that hierarchy of layers cannot be simplified directly. However a formal equivalence with a one-to-all direct connection might be possible (with the restriction of a lack of robustness of the "equivalent" simplified representation - see (Linsker, 1988)).

**Axiom 1** *In the case of an adapted<sup>5</sup> system, the action imposed at the output is equal to the sum of the input actions or currents. Because of the output non linearity, each output is able to provide the same output current at all the output links. The output is a source of current. Its value is computed from the input current but it is not the input current.*

Except for robustness and learning problems, a lot of simplifications can be introduced in a CS architecture so as to compare its functionality with another architecture. Of course, in an autonomous system with on-line learning, obtaining stable representations might be a problem. We postulate that for a given time interval the learned configurations will be stable enough so that some simplifications

<sup>5</sup>An adapted system is a system that can be considered as static in its structure and connectivity: the learning has stabilized its behavior. A more formal definition will be introduced in section 5 (definition 4).

can be performed (but only valid for this time interval).

## 4 Energy measure in a CS

At this point, the problem is that we have proposed relations between hypothetical intrinsic variables of a CS but that we don't have a relevant and useful notion of measure. It seems difficult to define a relevant intensive value associated to the vectorial input information. Perhaps something linked to the entropy of the vectors (like building contrast functions (Comon, 1996)) might be useful but we still have not found a simple way to use them. We will try to overcome this difficulty and deal directly with the energy that might be consumed by a CS according to its actions and their effects on the perception (we continue the analogy with electricity and physics).

**Definition 2** *The energy matrix  $E$  associated to a given Perception/Action system can be defined as the integral of  $Per.Ac^T$ :*

$$E(\tau) = \int |Per(t - \tau) \rangle \langle Ac(t)| dt \quad (7)$$

with  $\tau$  the input/output delay.

If the system is purely reactive, the energy taken from the outside to produce an action is  $E(\tau) = \Delta(\tau) \cdot G^T$  with

$$\Delta(\tau) = \int |Per(t - \tau) \rangle \langle Per(t)| dt$$

and  $G$  the conductance of the system.  $\Delta(\tau)$  can be seen as an auto-correlation function.

Obviously, an important bias can be introduced by very high component values in the perception and action vector. A normalization of the energy matrix must be performed. A very simple solution is the following:

$$E = \int \frac{|Per \rangle \langle Ac|}{\sqrt{Per^2 + Ac^2}} dt$$

This matricial representation of the energy can provide an intuitive understanding of the CS complexity. Fig. 15, 16, 17 and 18 show respectively simple examples of the matrix  $E$  associated to a constant system, a random system, a chemico-taxi behavior and a visual homing behavior.

**Definition 3** *The dimensionality of a behavior or its "Energy" dimensionality  $E_D$  (or its "complexity") can be expressed according to the rank of  $E$ .*

$$E_D = rank(E)$$

In the case of a constant action (fig. 15), it is trivial to show the matrix associated to the behavior has a single non-zero eigenvalue and the rank of the  $\int Per.Ac^T dt$  matrix is 1.

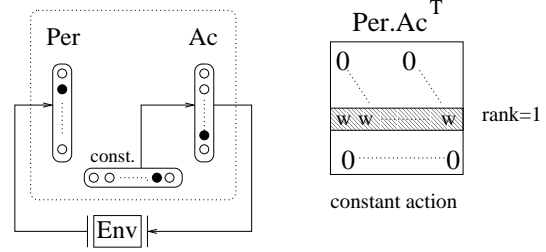


Figure 15: Representation of the  $\int Per.Ac^T dt$  matrix associated to a constant behavior (always the same action, whatever the perception is).

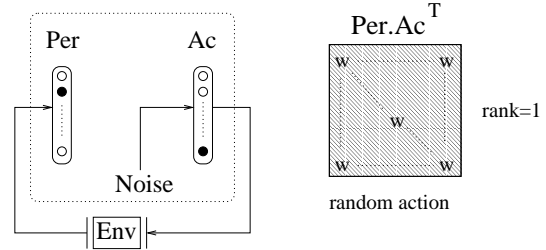


Figure 16: Representation of the  $\int Per.Ac^T dt$  matrix associated to a random system (no link between perception and action). The actions are selected randomly (they are independent from the perception). The matrix will tend to have a uniform value and all the eigenvalues excepted the first one can be neglected (see appendix).

The case of a random behavior is interesting since it can be either associated to a very low rank, 1, or to a full rank  $rank(E)=size(E)$ <sup>6</sup>. But in that last case (fig 16), the matrix coefficients will have almost the same value (because input and output are independent) and the matrix will tend to a rank 1 matrix (see appendix). So we can say the complexity of a random behavior and a constant behavior are the same and are the simplest behaviors different of a null behavior (rank=0).

The chemio-taxi behavior used by bacteria for instance can be considered as just a little bit more

<sup>6</sup>If the random matrix is rectangular, its size corresponds to its smallest dimension.



complex. The strategy is “go ahead” when the sugar concentration increases and moves randomly in all the other cases (fig. 17). The matrix rank  $E_D$  is equal to 2 (2 rectangular blocs of identical non null values).

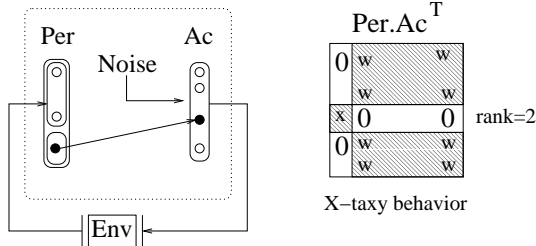


Figure 17: Representation of the  $\int Per.Ac^T dt$  matrix associated to a chimio-taxy behavior. The system moves randomly or turn constantly expected when the goal is in sight or a measure gradient increases. In this last case, the system goes straight ahead (bacteria chemio-taxy or simple photo-taxy behaviors could be explained this way).

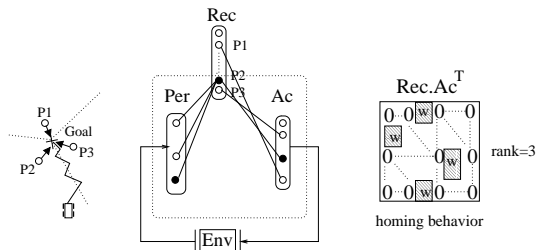


Figure 18: Representation of the  $\int Per.Ac^T dt$  matrix associated to a homing behavior. 3 places (P1, P2, P3) are recognized from the input perception flow and are associated to 3 actions. These Perception/Action associations create an attractor basin in which the system will descend.

A more complex behavior is the visual homing. It can be associated to a minimal value of  $E_D$  equal to 3. The simplified network supposes that it is possible to extract from the perceived flow, continuous enough information so that place recognition neurons will react over large continuous areas (see (Gaussier and Zrehen, 1995; Gaussier et al., 2000a) for more details). Hence, if 3 places are learned around a given goal and associated to the motor actions allowing to reach the goal, the competition in the place recognition group will allow the robot to move in the goal direction whatever its starting position in the visual field is.

To analyze more complex CS systems, the direct computation of  $E_D$  can be performed but another interesting solution would be to estimate  $E_D$  in

comparison to other CS. If we consider 2 CS,  $\psi_1$  and  $\psi_2$ , we can use the following properties:

$$E_D(\psi_1 + \psi_2) \leq E_D(\psi_1) + E_D(\psi_2)$$

$$E_D(\psi_1 \times \psi_2) \leq E_D(\psi_1) \times E_D(\psi_2)$$

$$E_D(\psi_1 \circ \psi_2) = MIN(E_D(\psi_1), E_D(\psi_2))$$

to bound the  $E_D$  value of any CS. These equations can also be used to characterize the simplicity or the cognitive cost of a local network in a given CS. We can consider the best solution, is the solution associated to the minimal rank of the energy matrix.

## 5 Analysis of a conditioning rule

Now, we will show the “energy” measure defined above can be used to retrieve the equations corresponding to the classical Pavlovian conditioning modeled by the Rescorla and Wagner learning rule (Rescorla and Wagner, 1972) in psychology and the Widrow and Hoff delta rule (Widrow and Hoff, 1960) in signal processing (these two equations are equivalent). The condition needed to deduce this rule from our energy measure depends on the adaptation definition.

**Definition 4 Adapted system** We will consider a system is adapted when there is no current or action lost at the level of the different groups.

Fig. 19 show a network for conditioning learning and its equivalent “electrical” circuit with the different current or action flows. To simplify the reasoning, the system is reduced to a 2 input neurons and 1 output neuron system.

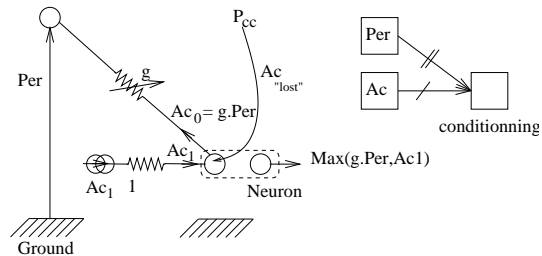


Figure 19: Equivalent electrical circuit of a conditioning mechanism.  $P_{cc}$  represents a constant source of energy. Learning consists in minimizing the consumption of this energy.

The associated equations for the Perception and Action flows are the following:  $Ac_0 = Ac_1 + Ac$  and

$g.Per = Ac_0$  which implies:

$$Ac = g.Per - Ac_1 \quad (8)$$

If we consider  $Ac$  as a current of lost, we can try to minimize its average value if we consider the input signal are stochastic variables. This can be written:

$$E[Ac^2] = E[(g.Per - Ac_1)^2] \quad (9)$$

The minimization of 9 in terms of  $g$  can be achieved by the classical LMS (least mean square) algorithm (Widrow and Hoff, 1960) or the Rescorla and Wagner rule in conditioning theory (Rescorla and Wagner, 1972). The learning rule is:

$$g(n+1) - g(n) = Per(n) (Per^T(n)g(n) - Ac_1(n))$$

Hence, in an adaptive filter, the action flow is represented by the error correction (the adaptation mechanism) according to the input data (the “perceived” information).

In the case of an unsupervised learning like a self organized map or a WTA mechanism, we can consider each neuron as a subsystem that tries to maximize the difference of potential between neurons because of the negative weights (equivalent to negative resistors). After adaptation the current lost should be minimal or null.

The problem with this kind of reasoning is that we have to come back to the description of each particular neuron model which is contradictory to a general analysis tool. It is clear that in a coherent theoretical frame, it should be possible to deduce the local adaptation rules from the global energy rule of the system  $\int |Per \rangle \langle Ac| dt$ . Unfortunately, if the neuron has one action input and one perception input, the product between them is a scalar and its maximization will induce an incorrect learning rule (no constraint on their difference of activity level). A more sophisticated measure could consist in discretizing the two inputs over vectors of size  $N$  (like a measure on a graduated rule).

For a scalar input  $x \in [0, 1]$ , the associated  $|x \rangle = (x_0, x_1, \dots, x_i, \dots, x_N)$  vector will be defined as follows:  $x_i = 1$  for  $i = \text{round}(x.N)$  and  $x_i = 0$  otherwise. We will call  $B$  this function:  $|x \rangle = B(x)$ . Now, if we consider the scalar input  $Ac$  and  $W.Per$  then

$$E = \int B(W.Per(t)) . B(Ac(t))^T dt \quad (10)$$

We want to maximize the rank of  $E$  ( $\text{rank}(E) = \text{dim}(E)$ ). If Action and Perception data are independent then the all the matrix values will tend to the same value and the matrix rank will be 1 (or a higher value if there are some residual structures in the input data). Of course, in the case the perception and the proposed action (desired action  $Ac_1$ ) are random independent signals, it will be impossible to increase the rank of the matrix and this agrees with the impossibility to learn something.

On the contrary, if there is a linear correlation between Perception and Action that the neuron can capture, then the matrix  $E$  will be null excepted for a line of non null values with a slope depending of the linear correlation factor and on the neuron weight (the rank of the matrix will depend on this slope see fig. 20).

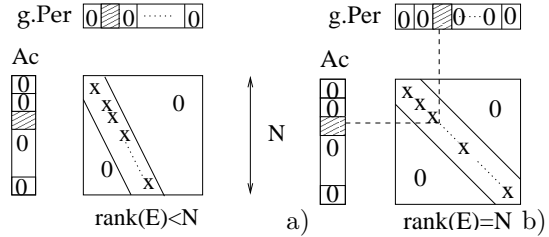


Figure 20: Energy matrix a) when the neuron does not predict  $Ac = W.Per$ . Here the weight  $W$  is too low. b) The energy matrix tends to be an identity matrix when  $Ac = W.Per$ .

To maximize the matrix  $E$  rank, we must have  $\lambda.Id = E$ . It corresponds to  $Ac(t) = W.Per(t)$ . The rank maximization is performed when the average error on  $Ac - W.Per$  is minimum. Fortunately, it is equivalent to the minimization of the current  $Ac$  associated to a loss of energy in the eq. 8 and will induce the same solution (see eq. 9).

## 6 Conclusions and perspectives

The theory outlined in this paper comes from the merging of two different ideas. First, there is the will to transform the philosophical ideas around the importance of the Perception/Action coupling in something more formal and systematic. Second, we observed in a brain modeling work on a structure called the hippocampus (Gaussier and Zrehen, 1995; Gaussier et al., 2001) that this structure and its associated cortical areas compute among other things the product between perception and action vectors ( $Per.Ac^T$ , see fig. 21) and has a short

term memory which could represent the integration mechanism. This Perception/Action product could be used by the brain to predict future Perception/Action transitions and to detect novelty. Hence the hippocampus could be seen as an apparatus able to measure a kind of cognitive energy consumption.

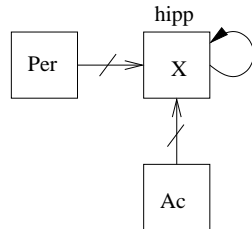


Figure 21: Simplified representation of the merging between Perception and Action performed at the input of the hippocampus. The result is an internal representation of the whole system/environment interactions (an operator to measure energy consumption).

Hence, it could be legitimate to represent Perception and Action as the two faces of the same coin. We can choose to speak about Perception or Action according to the fact data seem to represent more the perception or the action side but it is always the (Perception,Action) couple that will be manipulated as a single entity. The analogy with the electrical circuits is here very important to understand what we mean. For instance, it is sometimes easier to consider a current or a tension according to the circuit topology but at any place where a current exists, it is caused by a difference of tension, and reciprocally any mention induces a current (the same is true for Perception and Action).

Surprisingly, the analogy between the concept of Perception/Action in a CS and the effort/flow or tension/current in electricity was able to throw another light on a lot of important concepts in cognitive sciences such as the reactivity, the habituation, the sensibilization or the function of brain structure such as the hippocampus. Moreover, it seems possible with little modifications to apply the computation rules used in electricity on diagrams representing simple CS. The main difference is that each CS block is supposed to be connected to a virtual source of energy. This source must be able to provide "energy" in the case of a decorrelation or an inadaptation between its different input flows (incoherence between the different input perception or between the different proposed actions). This notion of virtual energy is interesting since it can be applied in the same way for biological cognitive systems and computer simulations.

At last, the capability to retrieve from these new considerations a classical and optimal learning rule seems to show there is an internal coherence in the proposed formalism. Obviously, this framework is really far away from a global and rigorous mathematical theory of CS but our goal was only to insist on the possibility and the interest of more formal analysis and comparisons between CS architectures. A good theory of CS should be able to :

- introduce some kind of abstract measures independent of the system specificities (like the notion of mass to describe any object in physics),
- propose simple writing rules able to describe any architecture
- foresee the main properties of a given architecture/environment couple,
- measure the adaptiveness of an architecture for a given couple of behavior/environment,
- compare and simplify different models or architectures,
- deal with different level of CS description (neurons or elementary logical or arithmetic operators, cell assemblies, functional blocs, or even social interactions between individuals),
- simplify the communication between researchers using different kind of tools to model CS or to control autonomous robots or agents.

In the present paper, we have tried to avoid measure problems by the only use of relative information between different sources of input and output. We have implicitly supposed there is no cost "in the cognitive sense" to transform an information of action into a real action. If it can be a valid hypothesis then we can certainly continue in that way and try to understand what is the "cost" of a particular adaptation of the behavior or the cost of the self-organization or re-organization of a map of neuron according to particular input and output. It is also clear that global reinforcement signal and essential internal variable of the system should have a strong and central place in a coherent theory of cognitive systems.

Future works will have to address the problem of the minimal set of building blocks necessary to describe any cognitive system. For instance, the wide literature on neural network modeling should be explored in order to find all the non-simplifiable structures. The reentrant maps introduced by Edelman (see fig. 22) are certainly a good example of

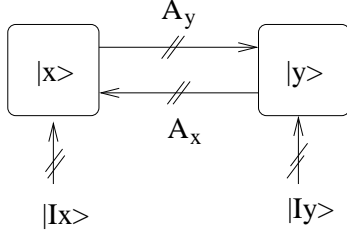


Figure 22: The reentrant maps proposed by Edelman

a structure that cannot be simplified. New mathematical formalisms have to be imagined to enlight the importance of the dynamics in these structures. There is certainly nowadays a place for a Cognitive System Algebra (CSA) useful to formalize the theoretical core of one science of the cognition (including the development of more “intelligent” autonomous robots).

## Appendix

### Random behavior analysis

**Theorem 1** *if  $|x \rangle$  and  $|y \rangle$  are 2 independent and random vectors, then the rank of  $E$  defined as follow  $E = \int |x \rangle \langle y| dt$  is 1.*

To illustrate this theorem, we compute the estimator:

$$\hat{E}(t) = \frac{1}{t} \sum_{i=0}^{t-1} |x_i \rangle \langle y_i| \quad (11)$$

on a long time interval and for random vector  $|x_i \rangle$  and  $|y_i \rangle$ . The results are presented fig. 23.

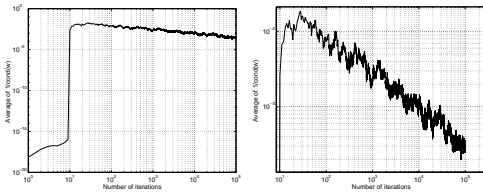


Figure 23: a) Evolution of the “energy” conditioning according to the number of integrated  $|Per \rangle \langle Ac|$  information. The scales are logarithmic to show the details of the beginning (matrix formation) and the long decrease of ratio which means the min eigenvector becomes more and more little in comparison to the maximal eigenvalue. The curve is an average of 30 realizations, with vectors of dimension 10. The vector components were drawn according to a uniform law  $[0,1]$ . b) details of the ratio decrease after the first 10th iterations.

Indeed, each component of the matrix will tend to the same value (central limit theorem) and then the rank of

the matrix will become 1. Since the values will never be exactly the same, the rank should theoretically remain full. In practice, one eigenvalue will be very high and all the other very low. We can consider a threshold and say that when the ratio between the max and min eigenvalues is too high then there is a “conditioning” problem (see fig. 23).

Other measures could be imagined to obtain more precise information on the behavior complexity but we can notice the present measure already allow to make some comparisons and justify the possibility to build a mathematical framework to study the mean features of cognitive systems.

### Example of a network simplification

The figure 24 represents an example of a network that can be simplified.

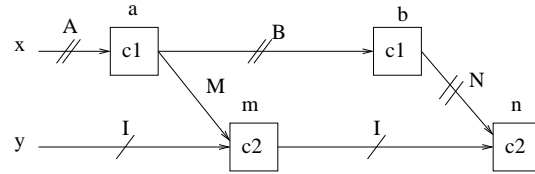


Figure 24: Example of a neural network that can be simplified. c1 and c2 represent 2 types of competitive structures. A, B, M and N represent the modifiable weights of the groups a,b,m and n respectively. I represents a “one to one” unmodifiable type of connection (reflex pathway).

Indeed, we have  $|b \rangle = c1|B|a \rangle$  and  $|n \rangle = c2|N|b \rangle$ . We can deduce  $|n \rangle = c2|N|c1|B|a \rangle$  that can be simplified in  $|n \rangle = c2|N'|a \rangle$  by using eq. 6. We obtain the equivalent network shown fig. 25.

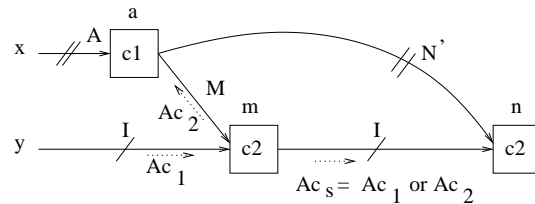


Figure 25: First simplification of the network fig. 24.

Now, if we consider y is an action flow (the one to one connections representing an action reflex pathway for instance) then after learning stabilization, we should not have current or Action lost between  $Ac_1 = y$  and  $Ac_2$  ( $Ac_s = Ac_1 \text{ or } Ac_2$ ). This implies the learning weights M and N are the same and  $|m \rangle = |n \rangle$ . The network fig. 24 is then equivalent to the very simple PerAc (Perception-Action) block that follows:

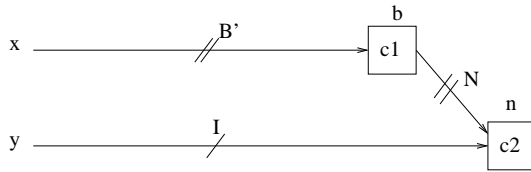


Figure 26: Simplified network equivalent to the network fig. 24.

## References

- Albus, J. (1991). Outline for a theory of intelligence. *IEEE trans. on syst. and cybern.*, 21(3):473–509.
- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87.
- Ashby, W. (1960). *Design for a brain*. London: Chapman and Hall.
- Berthouze, L. (2000). Bootstrapping the developmental process: the filter hypothesis. *Robot Learning - An Interdisciplinary Approach*, J. Demiris and A. Birk (eds), *World Scientific*, F68:8–30.
- Berthoz, A. (1997). *Le sens du mouvement*. Odile Jacob, Paris.
- Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, R.A. 2(1):14–23.
- Bugmann, G., Taylor, J., and Denham, M. (1995). Route finding by neural nets. In Taylor, J., editor, *Neural Networks*, pages 217–230, Henley-on-Thames. Alfred Waller Ltd.
- Chatila, R. (1995). Deliberation and reactivity in autonomous mobile robots. *Robotics and Autonomous System*, 16(2-4):197–211.
- Comon, P. (1996). Contrasts for multichannel blind deconvolution. *Signal Processing Letters*, 3(7):209–211.
- Daucé, E., Quoy, M., Cessac, B., Doyon, B., and Samuelides, M. (1998). Self-organization and pattern-induced reduction of dynamics in recurrent networks. *Neural Networks*, 11:521–533.
- Gaussier, P., Joulain, C., Banquet, J., Lepretre, S., and Revel, A. (2000a). The visual homing problem: an example of robotics/biology cross fertilization. *Robotics and autonomous system*, 30.
- Gaussier, P., Leprêtre, S., Quoy, M., Revel, A., Joulain, C., and Banquet, J. (2000b). Experiments and models about cognitive map learning for motivated navigation. In Demiris, J. and Birk, A., editors, *Interdisciplinary approaches to robot learning*, volume 24, pages 53–94. Robotics and Intelligent Systems Series, World Scientific, ISBN 981-02-4320-0.
- Gaussier, P., Revel, A., Banquet, J., and Babeau, V. (2001). From view cells and place cells to cognitive map learning: processing stages of the hippocampal system. *to appear in Biological Cybernetics*.
- Gaussier, P. and Zrehen, S. (1995). Perac: A neural architecture to control artificial animals. *Robotics and Autonomous System*, 16(2-4):291–320.
- Ikegami, T. (1993). Ecology of evolutionary game strategies. In *ECAL 93*, pages 527–536.
- Kelso, J. S. (1995). *Dynamic patterns: the self-organization of brain and behavior*. MIT Press.
- Kuniyoshi, Y. and Berthouze, L. (1998). Neural learning of embodied interaction dynamics. *Neural Networks*, 11(7/8):1259–1276.
- Linsker, R. (1988). Self-organization in a perceptual network. *IEEE Computer*, pages 105–115.
- Newell, A. (1982). The knowledge level. *Artificial Intelligence*, (18):87–127.
- Pfeifer, R. and Scheier, C. (1996). Sensory-motor coordination: the metaphor and beyond. *Robotics and Autonomous Systems*.
- Pfeifer, R. and Scheier, C. (1999). *Understanding intelligence*. MIT press.
- Quoy, M., Gaussier, P., Leprêtre, S., Revel, A., Joulain, C., and Banquet, J. (1999). A neural model for the visual navigation and planning of a mobile robot. In *Advances in Artificial Life, ECAL99*, volume 1674, pages 319–323.
- Rescorla, R. and Wagner, A. (1972). *Classical Conditioning II : Current Research and Theory*, chapter A theory of Pavlovian conditioning : Variations in the effectiveness of reinforcement and nonreinforcement. Appleton-Century-Crofts, New York.
- Rosenberg, R. and Karnopp, D. (1983). *Introduction to physical System Dynamics*. McGraw-Hill Book Compagny, New York.
- Schöner, G., Dose, M., and Engels, C. (1995). Dynamics of behavior: theory and applications for autonomous robot architectures. *Robotics and Autonomous System*, 16(2-4):213–245.
- Smithers, T. (1995). On quantitative performance measures of robot behaviour. *Robotics and Autonomous Systems*, 15:107–133.
- Steels, L. (1994). Mathematical analysis of behavior systems. In Gaussier, P. and Nicoud, J., editors, *From Perception to Action: PerAc'94*, pages 88–95. IEEE computer society press.
- Taylor, J. (1995). Modelling the mind by psyche (the european human brain project). In Fogelman-Soulié, F. and Gallinari, P., editors, *International Conference on Artificial Intelligence*, pages 543–548, Paris, France. EC2 and Cie.
- Tsuda, I. (2001). Towards an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *to appear as a Target Article in Behavioral and Brain Sciences*, 24(4).
- Widrow, B. and Hoff, M. E. (1960). Adaptive switching circuits. In *IRE WESCON*, pages 96–104, New York. Convention Record.
- Wiener, N. (1948, 1961). *CYBERNETICS or Control and Communication in the Animal and the Machine*. MIT Press.