Toward a Cognitive System Algebra: Application to facial expression learning and imitation

Philippe Gaussier[1], Ken Prepin[1,2], and Jacqueline Nadel[2]

[1] Neuro-cybernetic team, Image and Signal processing Lab., UMR CNRS 8051
Cergy Pontoise University / ENSEA, 6 av du Ponceau, 95014 Cergy , France
gaussier @ensea.fr
http://www.etis.ensea.fr/~neurocyber
[2] UMR CNRS 7593, Hopital la Pitié Salpétrière, Paris, France

Abstract. In this paper, we try to demonstrate the capability of a very simple architecture to learn to recognize and reproduce facial expressions without the innate capability to recognize the facial expressions of others. In the first part, the main properties of an algebra useful to describe architectures devoted to the control of autonomous and embodied "intelligent" systems are described. Next, we propose a very simple architecture and study the conditions for a stable behavior learning. We show the solution relies on the importance of the interactions with another system/agent knowing already a set of emotional expressions. A condition for the learning stability of the proposed architecture is derived. The teacher agent must act as a mirror of the baby agent (and not as a classical teacher). In conclusion, we discuss the limitations of the proposed formalism and encourage people to imagine more powerful theoretical frameworks in order to compare and analyze the different "intelligent" systems that could be developed.

1 Introduction

Nowadays hardware and software technologies allow to build more and more complex artifacts. Unfortunately, we are almost unable to compare two control architectures proposed to solve one given problem. Of course, one can try an experimental comparison on a given benchmark but the results focus on the optimality regarding the benchmark (how to deal with really unknown or unpredictable events?). We should be able to analyze, compare and predict in a formal way the behaviors of different control architectures. For instance, we must be able to decide if two architectures belong or not to the same family and can be reduced to a single architecture.

On another level, new design principles are proposed to create more "intelligent" systems [1] but there is no real formalization of these principles. The only way to correctly understand and use them is to have a long explanation build on examples showing cases of success stories (examples of good robotic architectures). Hence, we have good intuitions about

what to do or not to do to build a control architecture but it remains difficult to deal with really complex systems. Our situation can be compared to the period before Galileo when people knew objects fall but were unable to relate that to the concept of mass and acceleration in order to predict what will happen in new experiments. We urgently need tools to analyze both natural and artificial intelligent systems. Previous works have focused on mathematical tools to formalize pure behaviorist or reactive systems [2]. People have also tried with no real success to measure the complexity (in terms of fractal dimension for instance) of very simple behaviors like an obstacle avoidance [3]. The most interesting tools are dedicated to specific part of our global problem such as learning (see NN literature), dynamical systems [4] or some game theory aspects [5]. Yet, it remains difficult to overstep the old frame of the cybernetics [6, 7]. Finding the fundamental variables and parameters regarding some particular cognitive capabilities will be a a long and difficult work but we believe this should be related to the invariant properties of cognitive mechanisms and to the variation laws linking learning and embodiment.

In the present paper, we would like to show that a mathematical formalism used previously to represent for instance a control architecture dedicated to the visual homing [8], can also be used to build a simple theoretical model of the development of the capability to express and recognize more and more complex facial expressions. We will try to discuss, using this mathematical formalism, which are the basic mechanisms necessary to allow a naive agent to acquire the capability to understand/read the facial emotions of a teacher agent and to mimic them (so as to become a teacher and to allow turn taking in an emotion expression game). We will try to show that a newborn do not need a hardwired mechanism of emotion recognition to begin to interact in emotional games with adults. At last, we will discuss the drawback of the proposed formalism and try to propose directions for future researches since this work is at its very beginning.

2 Basic formalism of a Cognitive System

We summarize here the basis of our mathematical formalism. Figure 1 shows a typical control architecture for what we will call a cognitive [3] system (CS). The input and output of a CS are represented by vectors in the "bracket" notation[4]. An input or output vector x (column vector of size m) is noted $|x\rangle$ with $|x\rangle \in R^{+m}$ [5] while its transposed vector is

[3] The term cognitive must be understood here in the sense of the study of particular cognitive capabilities and not as a positive a priori for any kind of cognitivist approach.

[4] The formalism is inspired from Hilbert space used in quantum mechanics. Nevertheless, in our case it is not an Hilbert space since the operator is not linear...

[5] We consider the components of the different input/output vectors can only be positive/activated or null/inactivated. Negative activities are banned to avoid positive effects when combined with a negative weight matrix.

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 $n \times m$.

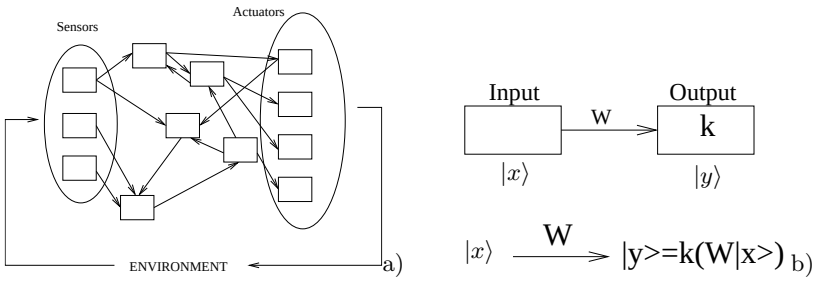

Fig. 1. a) Typical architecture that can be manipulated by our formalism. b) Graphical representation and formalism of the connection between 2 boxes in a CS.

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). We can consider that any element of a CS filters an input vector according to a matrix of weights W and a non-linear operator **k**. This operator represents the way to use the W matrix and the pattern of interactions between the elements of the same block. It can be a simple scalar product (or distance measure) or even a more complex operator such as an "If...then...else..." treatment (hard decision making), a pattern of lateral interactions in the case of a competitive structure, a recurrent feedback in the case of a dynamical system, a shifting mechanism, a mechanism to control a focus of the attention... Hence, we can consider these elements as "neurons" even if they can be more complex algorithmic elements in other programming languages. For instance, in the case of a simple WTA[6] box, we can write the WTA output $|y\rangle$ is $wta(A|x\rangle)$ with $|y\rangle = (0, ..., y_j, ...0)$ and $j = ArgMax(q_i)$ and $q_i = \langle A_i|x\rangle$. In the case of a Kohonen map, $|y\rangle = koh(A|x\rangle)$, the main difference is the way the output is computed: $q_i = \sum_j |A_{ij} - x_j|$. To be more precise, we should write $|y\rangle = koh(A, |x\rangle)$. Because, in the general case, an operator can have an arbitrary number of input groups, we will consider the recognition of an input is performed according to the type of its associated weight matrix. For instance, "one to one" input/output connections represented by the general identity weight matrix I is considered as the signature of a reflex pathway (because there is almost no interest to consider "one to one" learnable links). Basically, we distinguish 2 main types of connectivity according to their learning capabilities (learning possible or not): the "one to one" links (see fig. 2a) and the "one to many" connections (see fig. 2b) which are used for pattern matching processes, categorization... or all the other possible filtering. "One to many" connections will be represented in general by a A. In the

[6] Winner Takes All.

case of a complex competitive and conditioning structure with 1 unconditional (US) and 2 conditional (CS) inputs, we should write for instance $|y\rangle = c(A_1, |CS_1\rangle, A_2, |CS_2\rangle, I, |US\rangle)$. To avoid too many commas in the operator expression, we simply write $|y\rangle = c(A_1|CS_1\rangle, A_2|CS_2\rangle, I|US\rangle)$[7]. This allows to be sure a particular matrix is always associated to the correct input vector but it does not mean the matrix has to be multiplied by the vector (this computation choice is defined by the operator itself).

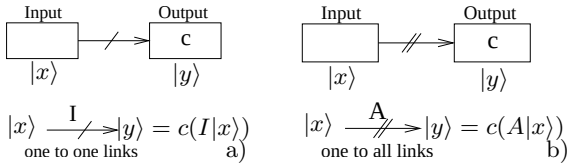

Fig. 2. Arrows with one stroke represent "one to one" reflex connections (one input connected to one output in an injective manner). Arrows with labels and 2 parallel strokes represent "one to many" modifiable connections between input and output nodes. a) Unconditional "one to one" connections (used as a reflex link) between two groups. Upper image is the graphical representation and lower image is the formal notation. b) "One to many" connections with a competitive group representing the categorization of the input stimulus at the level of the output group.

The main difference with classical automata networks is that most of our operators can adapt or learn online new input/output associations according to their associated learning rule. For instance, in the case of a classical Kohonen rule, we should write $\frac{dA_{ij}}{dt} = koh\_learning\,(|y\rangle, |x\rangle)$. Hence, 2 equations have to be written for each elementary box: one for the computation of the system output and another one for the weight adaptation (modification of the box memory). In the following, it will be crucial to remember our operators represent 2 different functions and flow of information moving in opposite directions. The first one will allow to transform sensorial information in an output code while the second one will act on the group memory in order to maintain a certain equilibrium defined by the learning rule [9].
In this paper, we will not discuss the interest or defaults of particular learning rules. Learning rules such as Least Mean Square algorithm (LMS - used for conditioning) or Hebb rule variants or any competitive rule will be sufficient for our demonstration since they are able to stabilize their associated weight matrices in the case the system is in a simple behavioral attractor or "perception state" (here simple means fixed point attractor).

Definition 1. *The perception Per can be seen as a scalar function ψ representing an attraction basin of the agent behavior. It can be seen as a sensori-motor invariant of the system (a kind of energy measure). Hence,*

[7] In previous papers, it was possible to write $|y\rangle = c(A_1|CS_1\rangle + A_2|CS_2\rangle + I|US\rangle)$ but many reviewers complained about the risk of misunderstanding the meaning of the operator $+$.

the perception can only be defined for an active system and is dependent of the system dynamical capabilities (kind of body, sensors and actuators). We will write: $Per(\boldsymbol{p}) = -\langle Ac|\boldsymbol{p}\rangle = -\int_{\boldsymbol{p}+\delta\boldsymbol{p}} Ac\, d\boldsymbol{r}$ *where* $|\boldsymbol{p}\rangle >$ *describes the position of the system in the considered environment*[8].

This corresponds to our intuition of the recognition as an attraction basin. We will say a system is in a stable state of perception if it is able to maintain itself in the associated attraction basin. Hence, learning to recognize an object (from visual, tactile, auditory... informations) can be seen as learning to maintain the system in a particular dynamical attraction basin [10]. More illustrations and justifications of this definition can be found in [11].
So when studying a control architecture, we will not need to take into account all the details of its implementation. We will have to focus on the global architecture and the way its elements are able to shape the behavior (building attractor basins).

3 Formal simplification rules

Now, the problem is to be able to simplify a CS architecture in another one (either simpler to analyze and to understand the architecture or more complex to provide more degrees of freedom to increase the architecture performances). Two architectures will be considered as equivalent if they have the same behavioral attractors (or perception state as defined previously). This means we cannot study a control architecture alone. The interactions with the environment must be taken into account. After the learning of a first behavior, the dynamics of the interactions with the environment (the perception state) is supposed to be stabilized. In the present formalism, two types of diagram simplifications will be considered. Simplifications of the first type can be performed at any time and leave the fundamental properties of the system completely unchanged (these are very restrictive simplification rules). Those of the second type only apply after learning stabilization (if learning is possible!). They allow strong simplifications but the resulting system is no more completely equivalent to the departure system (the new system will be less robust, less efficient and less precise for instance). At the opposite, the same formalism can be used to complexify an architecture in order to increase the efficiency of a given set of cognitive capabilities (increase of the system elasticity, robustness, precision...).
We present now a first example of simplification rule based on the existence of unconditional and reflex links. If we consider a linear chain of unconditional links between competitive structures of the same size such as "Winner Take All" (WTA), the intermediate competitive boxes are useless since they replicate on their output their input information. Hence we can write for instance that if we have: $|b\rangle = c(I|a\rangle)$ and $|d\rangle = c(I|b\rangle)$ then $|d\rangle = c(I|c(I|a\rangle))$ which should be equal to $|d\rangle = c(I|a\rangle)$ because a

[8] In naive cases, $|\boldsymbol{p}\rangle$ can be expressed in Cartesian coordinates or in any pertinent parameter space useful to describe more complex cases.

cascade of competitions leads to an isomorphism between the different output vectors which become equivalent to each other after the self organization of the different groups. So we can deduce the following rule $\mathbf{c}(\mathbf{I}|\mathbf{c}(.)) = \mathbf{c}(.)$. Other static simplification rules can be built in the same way [9]. Other simplifications can be used to represent the effect of learning. Except for robustness, these simplifications can be introduced to compare different control architectures (or to build more complex controllers). We will suppose that the system is in a stable state of perception or interaction with its environment. That is to say, it exists a time period where the system remains almost unchanged (internal modification must not have an effect on the system behavior). To be in a stable state, the environment properties must be quite constant. We postulate that for a given time interval, the learned configuration will be stable enough so that the simplifications can be applied (but they remain only valid for this time interval). Fig. 3 shows an intuitive representation of the evolution of a system behavior through time. The system behavior can

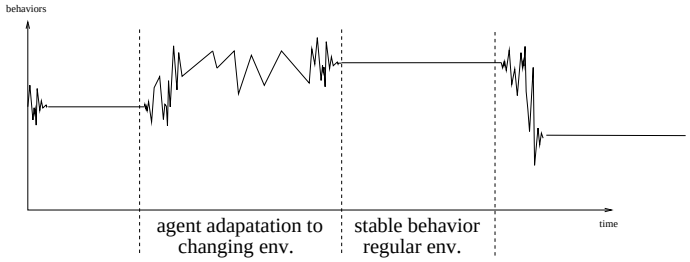

Fig. 3. Intuitive representation of what is a stable behavior allowing formal simplifications of the system.

evolve to adapt itself to an environment variation (or to the variation of an internal signal). In this case, it moves from a stable state to an unstable state or transition phase. It is only during the stable phases that the following simplifications can be considered as valid. Hence, we have to highlight a "before learning state" and an "after learning state" since some of the simplifications can be made at any time while some others must necessarily be made in the "after learning state".

A very simple example of such a simplification is the case of strict self organized learning group or competitive boxes (c operator) push-pully connected, fig. 4. We have $|y\rangle = c(A_1|x\rangle)$ and $|z\rangle = c(A_2|y\rangle)$ with A_1 and A_2 the matrices to learn the relevant input configurations. So $|z\rangle = \mathbf{c}\left(\mathbf{A_2}|\mathbf{c}\left(\mathbf{A_1}|\mathbf{x}\right)\right) = \mathbf{c}(\mathbf{A}|\mathbf{x}\rangle)$ since it is always possible to create a bijection between the activation of a given neuron in a first group and the activation of another neuron in a second group. Both sets of neurons can be considered as equivalents.

A more interesting case corresponds to the conditioning learning. The conditioning network (fig. 5 a) should be equivalent "after learning" to the simple network shown fig. 5 b and can be translated by the following

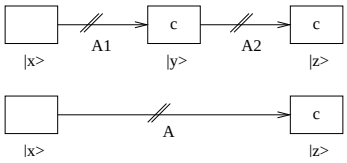

Fig. 4. A cascade of competitive or unsupervised classification structures can be simplified in a single competitive or classification box with a possible loss of performance but without a change in the main properties of the architecture.

equation: $c\left(I|US\rangle, A|CS\rangle\right) \approx c\left(A|CS\rangle\right)$ where $|US\rangle$ represents the unconditional stimulus and $|CS\rangle$ the conditional stimulus. The simplification "before learning" considers only the reflex pathway: $c\left(I|US\rangle, A|CS\rangle\right) \approx c(I|US\rangle)$ (functioning is equivalent in a short time delay but there is no possible adaptation) whereas the other simplification represents the equivalent NN in the "after learning" situation: not equivalent if the environment changes too much and leads the agent to be inadapted.

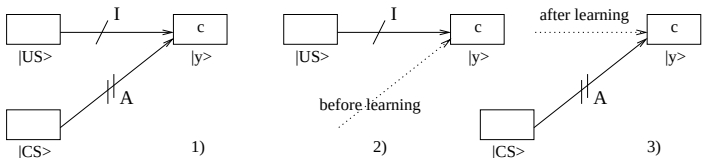

Fig. 5. Image 1 is the graphical representation of a conditioning learning $|y\rangle = c(I|US\rangle, A|CS\rangle)$. Image 2 is the graphical representation of the equivalent network before learning and Image 3 after learning $|y\rangle = c(A|CS\rangle)$.

We have shown in [9] that maximizing the dimensionality (rank) of the perception matrix $\sum_P |Ac\rangle\langle S|$ can be equivalent to the mean square error minimization performed when trying to optimize the conditioning learning between the action proposed by the conditional link and the action proposed by the unconditional link (where $|Ac\rangle$ represents the action vector (here $|y\rangle$) and $|S\rangle$ the sensorial input (here $|CS\rangle$)). Hence, learning can be seen as an optimization of the tensor representing the perception. In other words, we can say the proposed simplification rules are relevant if the system is adapted to its environment or if the system perceives its environment correctly according to the capabilities of its own control architecture (learning capabilities). We can notice that $Per = \sum_P \langle Ac|S\rangle = tr(\sum_P |Ac\rangle\langle S|)$ while the "complexity" of the system behavior can be estimated from $rank((\sum_P |Ac\rangle\langle S|)$.

4 Application to social interactions learning

In this section, our goal is to show how our formalism can be applied to analyze a very simple control architecture and justify some psycho-

logical models (see [12] for a discussion on the importance of an emotional system in autonomous agents). At the opposite to the classical pattern recognition approach, we will show that an online dynamical action/perception approach between two interacting systems has very important properties. The system we will consider is composed of two identical agents (same architecture) interacting in a neutral environment (see fig. 6).

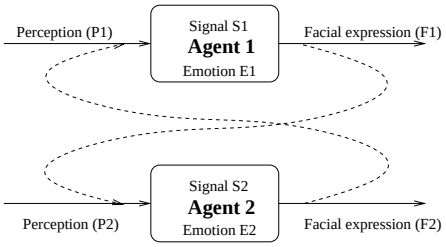

Fig. 6. The bidirectional dynamical system we are studying. Both agents face each other. Agent 1 is considered as a newborn and agent 2 as an adult mimicking the newborn facial expressions. Both agents are driven by internal signals which can induce the feeling of particular emotions.

One agent is defined as an adult with perfect emotion recognition capabilities and also the perfect capability to express an intentional emotion[9]. The second agent will be considered as a newborn without any previous learning on the social role of emotions. First, we will determine the conditions for a stable interaction and show that in this case learning to associate the recognition of a given facial expression with the agent own "emotions" is a behavioral attractor of the global system. Our agents receive some visual signals (P_i perception of agent i). They can learn and recognize them ($|R_i\rangle$ activity). Hence, the perception of a face displaying a particular expression should trigger the activation of a corresponding node in R_i. This mechanism can use an unsupervised pattern matching technics such as any winner take all mechanism (WTA, ART network, Kohonen map...).

$$|R_i\rangle = c\,(A_{i1}|P_i\rangle) \tag{1}$$

c represents a competitive mechanism allowing for instance to select a winner among all the vector components. To simplify, this winning component is put to 1 and the other ones to 0 (any other non linear competition rule could be applied and should not change our reasoning). A_{i1} represents the weights of the neurons in the recognition group of the agent i allowing a direct pattern matching. Our agents are also affected by the perception of their internal milieu (hunger, fear etc.). We will call S_i the internal signals linked to physiological inputs such as

[9] The problem of the dynamical development of two identical agents in a more free interaction game will be studied in a following paper.

fear, hunger... "Emotion" recognition E_i depends on the internal milieu. The recognition of a particular internal state will be called an emotional state E_i. We suppose also E_i depends on the visual recognition R_i of the visual signal P_i. At last, the agents can express a motor command F_i corresponding to a facial expression. If one agent can act as an adult, it must have the ability to "feel" the emotion recognized on someone else's face (empathy). At least, one connection between the visual recognition and the group of neuron representing its emotional state must exist. In order to display emotional state, we must also suppose there is a connection from the internal signals to the control of the facial expression. The connection can be direct or through another group devoted to the representation of emotions. For sake of homogeneity, we will consider that the internal signal activates through an unconditional link the emotion recognition group which activates through an unconditional connection the display of a facial expression (hence it is equivalent to a direct activation of F_i by S_i - see [9] for a formal analysis of this kind of properties). Hence, the sum of both flows of information can be formalized as follow:

$$|E_i\rangle = c\,(I|S_i\rangle, A_{i3}|R_i\rangle) \tag{2}$$

At last, we can also suppose the teacher agent can display a facial expression without "feeling" it (just by a mimicking behavior obtain form the recognition of the other facial expression). The motor output of the teacher facial expression then depends on both facial expression recognition and the will to express a particular emotion:

$$|F_i\rangle = c\,(I|E_i\rangle, A_{i2}|R_i\rangle) \tag{3}$$

Fig. 7 represents the network associated to the 3 previous equations describing our candidate architecture. In a more realistic architecture, some intermediate links allowing the inhibition of one or another pathway could be added but it is out of the scope of the present paper, which aims at illustrating what can be done with our formalism on a very simple example.

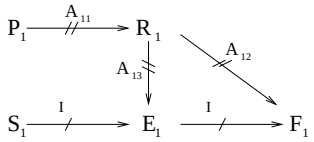

Fig. 7. Schematic representation of an agent that can display and recognize "emotions"(notations see fig. 2). Arrows with one stroke represent "one to one" reflex connections. Arrows with labels and 2 parallel strokes represent "one to all" modifiable connections.

4.1 Condition for learning stability

First, we can study the minimal conditions allowing the building of a global behavioral attractor (learning to imitate and to understand facial

expressions). Fig. 8 represents the complete system with both agents in interaction. It is considered as a virtual net that can be studied in the same way than an isolated architecture thus allowing to deal at the same time with the agent "intelligence" and with the effects of the embodiment and/or the dynamics of the action/perception loops.

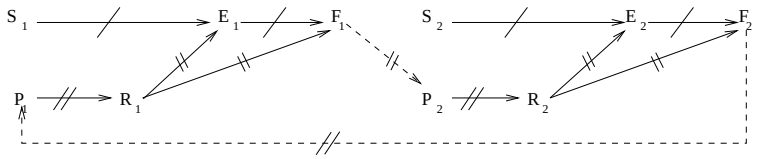

Fig. 8. Schematic representation of the global network representing the interaction between 2 identical emotional agents. The dashed links represent the connections from the display of a facial expression to the other agent perception system (effect of the environment).

The following simplifications apply before learning and concern only the unconditional links (see in the previous section the simplification of a conditioning structure before learning). We simply consider the activation of S can induce a reflex activation of a stereotyped facial expression F before (and after) the learning of the correct set of conditioning. The resulting network is shown fig. 9.

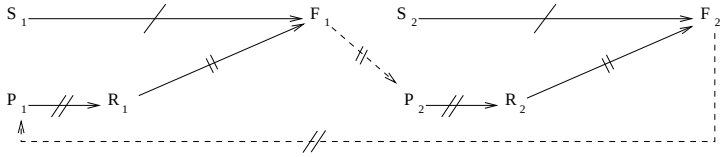

Fig. 9. Schematic representation of the simplified network representing the interaction between 2 identical emotional agents (modification of fig. 8)

Next, the linear chains of "one to many" modifiable connections and their associated competitive learning structures can also be simplified since $\mathbf{c}(\mathbf{A}|\mathbf{c}(.) \equiv \mathbf{c}(.)$. We finally obtain the network shown fig. 10 a).
It is much simpler on fig. 10 to see the condition of the learning stability. Since, the chosen simplifications allow to obtain a virtual network with learnable bidirectional connections between F_1 and F_2, a condition for the learning stability is that these connection weights remain stable. If S_1 and S_2 are independent, learning cannot be stable since S_1 and S_2 are connected through unconditional links to F_1 and F_2 respectively. The only way to stabilize learning is to suppose S_1 and S_2 are congruent. Otherwise a lot of "energy" is lost to adapt continuously the connections between F_1 and F_2 (see [9] for more details). Because, the agent rep-

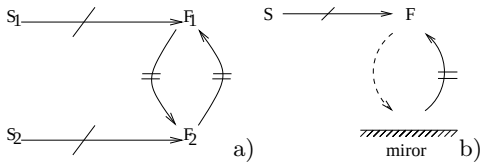

Fig. 10. a) Final simplification of the network representing the interaction between 2 identical emotional agents (modification of fig. 9). b) Minimal architecture allowing the agent to learn "internal state"-"facial expression" associations.

resenting the baby must not be explicitly supervised, a simple solution is to consider the agent representing the parent is nothing more than a mirror[10]. We obtain the network shown in fig. 10 b) where the architecture allows the system to learn the "internal state"-"facial expression" associations. Hence, we show that from our initial control architecture, learning is only possible if the teacher/parent agent imitates the baby agent. The roles are switched according to the classical point of view of AI and learning theory. This shows how taking account the dynamics of interactions between two agents can change our way of thinking learning and more generally cognition problems.

4.2 Learning the emotional value of facial expressions

These first simplifications bring us to the conclusion that learning stabilization is possible if the teacher/parent agent acts as an imitator of the baby agent. Now, we will suppose these conditions are respected. From the initial equations of the system, we will derive another set of simplifications in order to prove the beginner (or naive) agent can learn to associate the visual facial expression displayed by the teacher agent to the correct emotional state. We suppose the agent 1 perceptive input P_1 is the result of a linear projection of the facial expression (output) of the agent 2 and vice versa. We will write $|P_1\rangle = B_1|F_2\rangle$ and $P_2 = B_2|F_1\rangle$. Hence, $|R_1\rangle = c\,(A_{11}|P_1\rangle) = c\,(A_{11}.B_1|F_2\rangle) = c\,(A'_{11}|F_2\rangle)$ (with $A'_{11} = A_{11}.B_1$). We can then replace in this new expression of R_1, $|F_2\rangle$ by the result of the computation of the second agent (using eq. 3). We obtain:

$$\begin{aligned} |R_1\rangle &= c\left(A'_{11}|c\,(I|E_2\rangle, A_{23}|R_2\rangle)\right) \\ &= c\left(A'_{11}|c\,(I|E_2\rangle, A_{23}|c\,(A_{21}|P_2\rangle)\right)) \end{aligned}$$

[10] Obviously, another possible solution is that the second agent tries to deceive the first agent. If the second agent displays an "unhappy face" every time the first agent displays an "happy face" and vice versa an incorrect learning is possible. Fortunately, the probability of such a learning is very low if the first agent interacts with several independent agents (no conspiracy!). Yet, we can predict that a baby interacting with a depressed mother (low probability of "happy face") will have some difficulties to create an unbiased repertory for the recognition of other's emotional states.

On the other side, we have $|P_2\rangle = B_2|F_1\rangle$ so:

$$\begin{aligned} |R_1\rangle &= c\left(A'_{11}|c\left(I|E_2\rangle, A_{23}|c\left(A_{21}\cdot B_2|F_1\rangle\right)\right)\right) \\ &= c\left(A'_{11}|c\left(I|E_2\rangle, A_{23}|c\left(A'_{21}|F_1\rangle\right)\right)\right) \end{aligned} \tag{4}$$

A'_{21} is defined as the matrix resulting from $A_{21}\cdot B_2$. The equation 4 can be represented by the virtual[11] network shown fig. 11. Intuitively, the network means the visual recognition in the first agent depends on the emotional state of the second agent and should also be a function of the facial expression of agent 1.

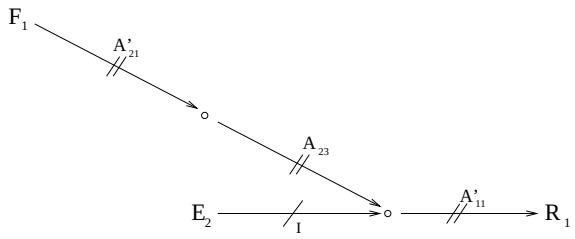

Fig. 11. Virtual net associated to eq. 4.

All the preceding simplifications could be made at any time (here, it is before learning). The following simplification can be done only after learning (and need the learning stability condition i.e. the second agent is a mirror of the first one). If the obtention of learning is possible (the error between F_1 and E_2 can be minimized in the mean square sense), conditioning learning in eq. 4 should result in:

$$I|E_2\rangle \approx A_{23}.c\left(A'_{21}|F_1\rangle\right) \tag{5}$$

if both architectures are identical, since there is no influence of learning on this simplification, we obtain by symmetry:

$$|E_1\rangle \approx A_{13}.c\left(A'_{11}|F_2\rangle\right) \tag{6}$$

Then, we can simplify eq. 4.

$$\begin{aligned} |R_1\rangle &\approx c\left(A'_{11}|c\left(A_{23}|c\left(A'_{21}|F_1\rangle\right)\right)\right) \\ &\approx c\left(A'_{123}|F_1\rangle\right) \end{aligned} \tag{7}$$

(we also have $|R_1\rangle \approx c\left(A'_{12}|E_2\rangle\right)$ but we won't use it.) Eq. 7 can be interpreted as the fact the activity of agent 1 visual face recognition is a function of its own facial expression. If we replace the value of F_1 obtained from eq. 3 in eq. 7, we obtain:

$$|R_1\rangle \approx c\left(A'_{123}|c\left(I|E_1\rangle, A_{13}|R_1\rangle\right)\right) \tag{8}$$

[11] This network is virtual since it mixes together parts of networks belonging to two different agents.

Here again $I|E_1\rangle$ is the reflex link and $A_{13}|R_1\rangle$ the conditional information. The conditional link can learn to provide the same results as the reflex link. If E_1 can be associated to R_1 then we obtain:

$$\begin{aligned} |R_1\rangle &\approx c\left(A'_{123}|c\left(I|E_1\rangle\right)\right) \\ \text{and } |R_1\rangle &\approx c\left(A'_{123}|E_1\rangle\right) \end{aligned} \tag{9}$$

This result shows the activity of the face recognition system is a direct function of the agent emotional state (R_1 can be deduce from E_1). In conjunction with the relation linking E_1 to R_1 (eq. 2) we can deduce the agent 1 (baby) has learned to associate the visual recognition of the tested facial expressions to its own internal feeling (E_1). The agent has learned how to connect the felt but unseen movements of self with the seen but unfelt movements of the other. It could be generalized to other movements since we showed in [13, 14, 15] that a simple sensori-motor system is sufficient to trigger low level imitations.

5 Discussion and perspectives

In this paper, we have applied a formalism proposed in [9] to simplify an "intelligent" system and to analyze some of its properties. We have shown a very simple architecture can learn the bidirectional association between an internal "emotion" and its associated facial expression. To demonstrate this feature, we have proved first that learning is only possible if one of the agents acts as a mirror of another. We have proposed a theoretical model that can be used as a tool not only to understand artificial emotional brains but also natural emotional brains. Let us consider a newborn. She expresses internal states of pleasure, discomfort, disgust, etc, but she is not aware of what she expresses. Within our theoretical framework, we can expect that she will learn main associations between what she expresses and what she experiences through her partners' mirroring of her own expressions. Seeing what she feels will allow the infant to associate her internal state with an external signal (i.e. her facial expression mirrored by someone else). Empirical studies of mother-infant communication support this view. For instance, two-month-old infants facing a non contingent televised mother who mirrors their facial expressions with a delay become wary, show discomfort and stop imitating the mother's facial expressions (see [16]). The primary need of mirroring is also demonstrated by the progressive disappearance of facial expressions in infants born blind. Another prospective benefit of the model is to give a simple developmental explanation of how facial expressions come to inform the growing infant about external events through the facial reading of what those events trigger in others [17]. Finally the model leads to suggest a main distinction between two processes of emotional matching: matching a facial emotion without sharing the emotion expressed: in this case there is a decoupling (see [18]) between what is felt and what is shown, thus it is pure imitation, and matching a facial emotion with emotional sharing, that is to say feeling what the other expresses through the process of mirroring, a definition of empathy (see [19]). More

complex architectures could be built on the basis of the studied model. For instance, adding feedback connections from the proprioceptive signal linked to the control of the facial expressions onto the recognition of an internal emotional state would allow the agent to "feel" a little bit more happy when he is smiling. Fig. 12 shows what could be such an architecture.

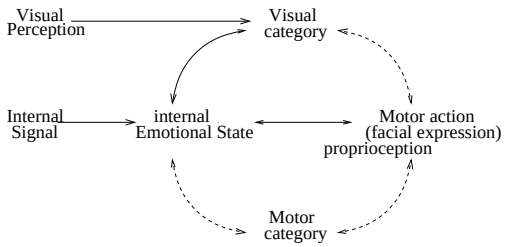

Fig. 12. Schematic representation of an agent that can show and recognize emotional states with feedbacks connections from action/proprioception to the different internal categorization structures.

In our lab., the formalism developed in this paper is used as a programming language allowing to describe architectures for visual object recognition, visual navigation, planning, visuo-motor control of an arm and imitation games... Yet, even if the size of our networks becomes more and more important, their intrinsic complexity remains low since it was our goal to prove complex dynamical behaviors could emerge from relatively simple architectures. At the opposite, Sporns et al. [20] study complex networks in terms of their structure and dynamics. They show highly complex networks have distinct structural characteristics such as clustered connectivity and short wiring length similar to those of large-scale networks of the cerebral cortex. Hence, future works will have to answer the following questions: Which are the minimal structures that cannot be simplified (or which intrinsic property is lost when an important simplification is made)? Which kind of really different operators have to be considered? And at a higher level, we will have also to understand how to manage different learning time constants and how to represent the body/controller codevelopment.

To answer these questions, it becomes necessary to test our formalism on other architectures and other problems to analyze precisely its limitations and to propose a more powerful framework. Another problem is that the demonstration proposed in this paper was based on the fact it was possible to isolate fixed point dynamics and to study one of them isolatedly. For more complex dynamical systems, we believe the same approach could be used (i.e. isolate the different dynamical regime and propose simplifications for each of them). Nevertheless, we believe the proposed approach could be directly applied to problems of the same kind of complexity level such as the problem of joint attention learning (see for instance [21] where a robot controller based on a principle close

to the one developed in our own architecture is proposed).

To sum up, we have shown it is possible to develop theoretical tools taking into account the interactions with the environment in order to compare and analyze different control architectures. In this context, the more a system is embodied, the less it need explicit learning since it is well adapted to its function: it relies on the physical plasticity of its physical architecture (see for instance how our mechanical anatomy simplifies the wide variety of tasks we have to solve). The need for learning can be seen as the impossibility of a perfect embodiment according to a given ecological niche (need of a physical compromise between the requirement of the different behaviors). Hence, in an algebra of embodied cognitive systems, we will have to distinguish between 3 levels of cognition: an infra level of cognition linked to the physical properties of the body, the individual level of cognition (the control architecture for one isolated agent) and the social level dealing with the social interactions between agents. At each level, the measure of the system elasticity or adaptation capability might be performed to characterize the embodiment of the system (see [9] for a tool to compare the complexity of different control architectures in term of an energy measure). In conclusion, we believe the difficulty of a formal analysis of cognitive systems is much more a problem of choosing the correct postulates and axioms than the lack of mathematical tools to deal with the intrinsic complexity of the existing systems.

Acknowledgements: This work is supported by the CNRS "ACI Computational Neurosciences" and "ACI Time and Brain" and CNRS team project on "imitation in robotics and development".

References

[1] Pfeifer, R., Scheier, C.: Understanding intelligence. MIT press (1999)
[2] Steels, L.: A case study in the behavior-oriented design of autonomous agents. In: SAB'94. (1994) 445–451
[3] Smithers, T.: On quantitative performance measures of robot behaviour. Robotics and Autonomous Systems **15** (1995) 107–133
[4] Schöner, G., Dose, M., Engels, C.: Dynamics of behavior: theory and applications for autonomous robot architectures. Robotics and Autonomous System **16** (1995) 213–245
[5] Ikegami, T.: Ecology of evolutionary game strategies. In: ECAL 93. (1993) 527–536
[6] Wiener, N.: CYBERNETICS or Control and Communication in the Animal and the Machine. MIT Press (1948, 1961)
[7] Ashby, W.: Design for a brain. London: Chapman and Hall (1960)
[8] Gaussier, P., Zrehen, S.: Perac: A neural architecture to control artificial animals. Robotics and Autonomous System **16** (1995) 291–320

[9] Gaussier, P.: Toward a cognitive system algebra: A perception/action perspective. In: European Workshop on Learning Robots (EWLR), http://www-etis.ensea.fr/~neurocyber /EWRL2001_gaussier.pdf (2001) 88–100
[10] Gibson, J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)
[11] Gaussier, P., Baccon, J., Prepin, K., Nadel, J., Hafemeister, L.: Formalization of recognition, affordances and learning in isolated or interacting animats. In: to appear in SAB04 (From Animal to Animat). (2004)
[12] Canamero, L.: Emotions and adaptation in autonomous agents: A design perspective. Cybernetics and Systems **32** (2001) 507–529
[13] Gaussier, P., Moga, S., Quoy, M., Banquet, J.: From perception-action loops to imitation processes: a bottom-up approach of learning by imitation. Applied Artificial Intelligence **12** (1998) 701–727
[14] Andry, P., Gaussier, P., Moga, S., Banquet, J., Nadel, J.: Learning and communication in imitation: An autonomous robot perspective. IEEE transactions on Systems, Man and Cybernetics, Part A **31** (2001) 431–444
[15] Andry, P., P.Gaussier, Nadel, J.: From sensorimotor coordination to low level imitation. In: Second international workshop on epigenetic robotics. (2002) 7–15
[16] Nadel, J., Revel, A., Andry, P., Gaussier, P.: Toward communication: first imitations in infants, low-functioning children with autism and robots. Interaction Studies **5** (2004) 45–75
[17] Feinman, S.: Social referencing and the social construction of the reality in infancy. Plenum Press, New York (1992)
[18] Scherer, K.: Emotion as a multicomponent process. Rev. Person. Soc. Psychol. **5** (1984) 37–63
[19] Decety, J., Chaminade, T.: Neural correlates of feeling sympathy. Neuropsychologia **42** (2002) 127–138
[20] Sporns, O., Tononi, G.: Classes of networks connectivity and dynamics. Complexity **7** (2002) 28–38
[21] Nagai, Y., Hosoda, Asada, M.: How does an infant acquire the ability of joint attention?: A constructive approach. In: Proceedings of the 3rd International Workshop on Epigenetic Robotics. (2003) 91–98